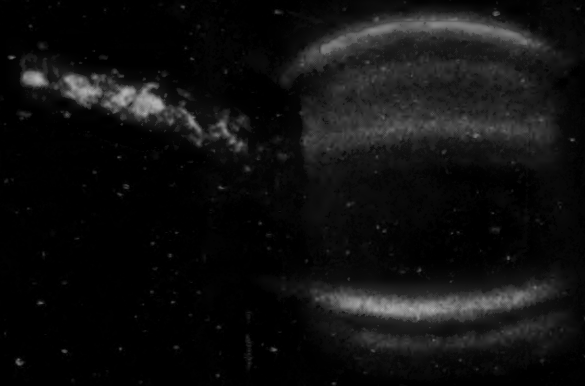
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 1 JANUARY, 1951

In This Issue

		Page
Theory of the Dynamic Plastic Deformation of a Thin Diaphragm	G. E. Hudson	1
Improvements in Visual Depiction of Magnetic Lines of Force by Means of a Gas Discharge	F. Blaha and J. A. Schedling	11
Electric Dipoles in the Presence of Elliptic and Circular Cylinders	Winston S. Lucke	14
Applications of Potential Theory to the Design of Linear Arrays	Thomas T. Taylor and John R. Whinnery	19
Theory of the Circular Diffraction Antenna	Harold Levine and Charles H. Papas	29
Digital Information Storage in Three Dimensions Using Magnetic Cores	Jay W. Forrester	44
Erratum: Erratum to Mechanical Properties of Glass	F. W. Preston	48
Parametric Excitation	N. Minorsky	49
Effective Earth's Radius for Radiowave Propagation Beyond the Horizon	William Miller	55
Experimental Measurement of Incompressible Flow Along a Cylinder with a Conical Nose	E. V. Laitone	63
Graphical Determination of the Fermi Level in a Simple Impurity Semiconductor	K. Lehovec and H. Kedesdy	65
Expandability of a Wave-Guide Field in Terms of Normal Modes	J. Van Bladel	68
Certain Refractory Compounds as Thermionic Emitters	D. L. Goldwater and R. E. Haddad	70
Calculation of Diffusion Penetration Curves for Surface and Grain Boundary Diffusion	J. C. Fisher	74
Secondary Emission of Nickel-Barium Mixtures and Rhenium When Bombarded by Electrons with Energies from 50 to 8000 Electron Volts	H. E. Farnsworth and M. J. Lun	77
Deterioration of Oxide-Coated Cathodes under Low Duty-Factor Operation	J. F. Waymouth, Jr.	80
Associated Legendre Polynomial Approximations	R. Landauer	87
Pulse Method of Determining the Energy Distribution of Secondary Electrons from Insulators	Kenneth G. McKay	89
Microwave Measurements of the Dielectric Properties of Gases	George Birnbaum, S. J. Kryder, and Harold Lyons	95
Books Received		94
Letters to the Editor:		
Note on the Metallic Behavior of $MoSi_2$	Frank W. Glaser	103
Comparison of Quenching Distance and Inflammability Limit Data for Propane-Air Flames	Dorothy Martin Simon	103
Note on Static Electrification of Dust Particles on Dispersion into a Cloud	John E. Coolidge and George Schulz	103
Comment on Static Electrification of Dust Particles on Dispersion into a Cloud	Wulf B. Kunkel	103
Condenser Aperture Device for Electron Microscopes	J. D. Boadway	104
Plasticity and Conductivity, Analogous Flow Phenomena in Copper Alloys	Rolland Sydney French	105
Noncontact Temperature Measuring Device	J. E. Field and S. D. Gehman	106
Magnetic Scaling Circuit	Hellmuth Hertz	107
Erratum: Theory of the Parallel Plane Diode	A. H. Taub and Nelson Wax	108
Spectral Emissivity of Coatings of Thoria and Other Refractories as a Function of Temperature	F. H. Morgan	108
Natural Width of a Grain Boundary in a Metal	J. L. Snoek	109
Program of the Electron Microscope Society of America		110
Index to Advertisers		xxvi

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
Lyle D. Borst
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
J. B. Fisk
G. P. Harnwell
George R. Harrison
W. V. Houston
Rudolph Kingslake
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager and Assistant Treasurer*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Senior Accountant*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Physics Lib.
Faxon

Journal
of
Applied Physics

Volume 22, Number 1 January, 1951

A Theory of the Dynamic Plastic Deformation of a Thin Diaphragm*

G. E. Hudson
Physics Department, College of Engineering, New York University, New York, New York
(Received February 23, 1950)

The theory presented in this article was developed in an attempt to describe the observed motion and plastic deformation of clamped metal diaphragms used in certain underwater explosion experiments and in certain mechanical gauges. The theoretical attack on this problem enables one to set up certain equations of motion, which may be solved in finite form under certain conditions. The solutions enable one to specify, for instance, the final deformed diaphragm profile, the distribution of thickness after deformation, the swing-time, which is the total time for deformation to take place, and many other quantities.

The simplest case, termed the "elementary approximation," turns out, except for relatively minor details, to describe adequately for many purposes the motion and final shape of the diaphragm; it is found that the deformed diaphragm shape is conical, the thickness distribution shows a marked dimpling at the center of the diaphragm, and the swing-time t_S is, to this order of approximation

$$t_S = a/c,$$

where a is the radius of the diaphragm, and c is the square root of the ratio of the "yield stress" to the density. These results are all in good agreement with experimental facts.

HISTORICAL BACKGROUND

IN connection with the torpedo-protection program which was conducted at the Taylor Model Basin, U. S. Navy, during the War, detailed studies were made of the damage done by underwater explosions to thin metal circular diaphragms, air-backed, and held rigidly at their peripheries. The idea was, that if an adequate description and explanation of the phenomena attendant on the damage to such an apparently simple structure could be obtained, some progress toward an understanding of damage produced similarly in more complex structures might be made. At the same time smaller metal diaphragms were used by several other research groups as underwater gauges for estimating from their deformations the relative strengths of underwater explosions. These investigations led to certain theoretical developments along this line but from a point of view different from that taken in this article.† On a later occasion, the necessity arose for the development of a simple mechanical gauge to measure the velocities acquired impulsively by large structures when they are subjected to a very great force of very brief duration. For this purpose, a small thin lead diaphragm was mounted in a rigid closed container which was then attached to a given structure. When an impulsive velocity was imparted to the structure, normal to the plane of the diaphragm, the diaphragm container moved and the diaphragm material tended to remain behind. The resulting deformation was of course the same as if the diaphragm itself had suddenly had impressed on it an equal but opposite uniform velocity normal to its plane while the container remained fixed, as in Fig. 1.

* This article is a revision and extension of a report written while the writer was at the David Taylor Model Basin, U. S. Navy. It was then entitled "A Theory of the Impulsive Plastic Motion of a Thin Diaphragm Normal to its Initial Plane" and was Part 18 of TMB Report R-254 (June 1944). Part of this paper was the subject of short talks given at the American Physical Society Meeting (Fluid Dynamics Section), Washington, D. C., May 1947, and at the Symposium on Plasticity, Brown University, Providence, Rhode Island, 9 February 1948. It is expected that the article will be one paper of a Compendium on British and American work on underwater explosion research to be published under the auspices of the U. S. Navy, with Dr. G. K. Hartmann, Chief of the Explosives Research Department, Naval Ordnance Laboratory, White Oak, Maryland, as American Editor, and Dr. E. G. Hill, Chief Scientist, Naval Construction Research Establishment, Rosyth, Scotland, as British Editor.

† See, for example, J. G. Kirkwood and J. M. Richardson, "The plastic deformation of circular diaphragms under dynamic loading by an underwater explosion wave," OSRD 4200 (1944). This unclassified report is a summary of most of the work presented in two earlier reports by Kirkwood, OSRD 793, and 1115, also declassified.

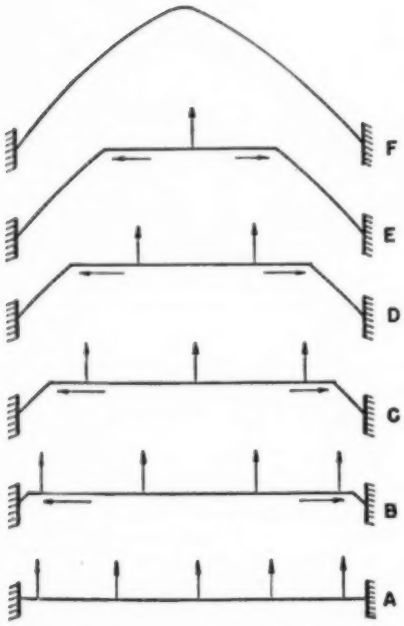


FIG. 1. (A) The diaphragm receives an initial impulsive velocity v, denoted by vertical arrows. (B) The material has a radial velocity distribution, denoted by horizontal arrows, and is restrained from moving at the periphery. (C) The bending wave, shown by the sharp corners in the profile has progressed inward from the edge. (D) As the diaphragm deforms further, the central region remains flat and moves with its initial velocity v. The bending wave is traveling into the central region and progressively deforming it. (E) As the bending wave nears the center it speeds up slightly. (F) The speeding up of the bending wave results in rounding off the apex of the conically deformed diaphragm.

The theory presented in this article was developed in an attempt to describe the observed motion and deformation of metal diaphragms under some of the conditions encountered in these experiments and with these instruments.

BASIC SUPPOSITIONS

The exact nonlinear partial differential equations of motion describing the dynamic plastic deformation of the metal in a diaphragm such as we have been considering are extremely complex. Even if it were possible to solve these with existing mathematical techniques the solution would without doubt be so unwieldy as to necessitate the introduction of radical simplifying approximations. Hence, it seems desirable to attack the theoretical problem of analysis of this motion by the artifice of replacing the actual mechanical system by a fictitious one in which are incorporated certain constraints which do no work and which operate to simplify the equations of motion. At the same time, this idealized model must be so chosen that most of the main qualitative features of the diaphragm motion, as actually observed, are preserved. In this way, it is hoped that the motion of the model when found agrees closely enough with that of the real system so that certain quantities, such as thickness distribution, central deflection, total strain, and time of deformation do not differ too much in magnitude from their actually observed values.

Consider, therefore, an ideal thin metal circular diaphragm of uniform thickness h, and radius a, held rigidly at its periphery. Initially, suppose that the diaphragm material has a uniform velocity component v normal to its initial plane. Because in the actual diaphragm, an elastic wave may then quickly set the material in motion radially, we shall suppose in addition that in the ideal diaphragm, there may be an initial linear radial velocity distribution superimposed on the uniform normal velocity v.‡ Estimates of the magnitude of this effect will be made in a later section.

At any later instant, the situation is considered to be as follows. A plastic bending wave has traveled inward some distance from the edge. In connection with this wave, we shall suppose that its shape is as shown in Fig. 2; that is, the bending wave represents a true discontinuity in the slope of the instantaneous profile of the diaphragm. As the wave sweeps over each annular material element in the flat central region, in effect it tilts it into the shape of an annular truncated conical element behind the wave. During this process, it is assumed that, because of the thinness of the diaphragm, no significant amount of work is done in bending. Moreover, no impulsive thinning is supposed to take place in the bend of the wave. We further suppose that the only stresses of importance at the bending wave are radial and circumferential principal stresses just ahead of the wave, and behind the wave in the tilted region, a normal stress component along the generator of the tilted element, a shear stress component, and a circumferential normal stress component.

Now the material passed over by the wave shall be supposed to have come to rest. The stresses which exist in this region are supposed to be such as to just hold the material in equilibrium—although their tendency to

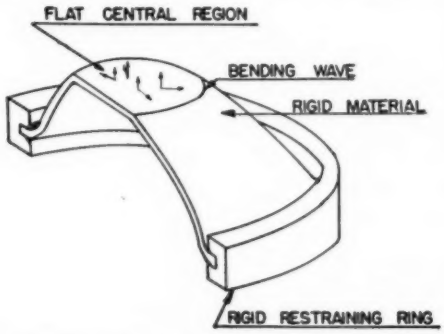


FIG. 2. Cut-away sketch of the diaphragm at some instant after the motion has begun. The flat central plastic flow region ahead of the bending wave, the wave itself, and the deformed region behind the bending wave, in which the material is supposed to be rigid and at rest, are clearly shown.

‡ This initial elastic stress phase has been investigated, but not published to the writer's knowledge, by F. Bohnenblust. He indicated orally some years ago to the writer that the circumferential stress component rises very quickly behind the elastic-stress wave in which the only nonvanishing stress component is the principal radial stress.

cause further plastic flow or not is a subject for further investigation, and might constitute a partial theoretical test of the model.

Within the central region which is as yet unaffected by the plastic bending wave, the diaphragm is still supposed to be flat, and therefore in the absence of normal force components it retains its uniform normal velocity v. At the same time, the material in the flat central region is supposed to be flowing radially outward and thinning. To avoid the extreme complexities of considering the probably nonuniform distributions of tension and thinning in an actual diaphragm, we introduce in this region a distribution of constraint forces which do no work and which produce a uniform thinning over the region during the motion. These constraints must be considered to yield a nonuniform stress system which is superimposed on the stresses arising from the uniform plastic deformation of this region.§ Their total effect is simply to bring about a simple radial velocity distribution and to afford a means of transferring the kinetic energy lost at the bending wave into kinetic energy of radial motion in the flat central region. Kinetic energy of thinning is supposedly very tiny, and is neglected.

As regards the properties of the metal of which the diaphragm is composed, we shall suppose the following. Because we are here dealing with finite deformations and strains, all purely elastic effects, other than those of the initial stress wave which have already been mentioned, would seem to be negligible. Consequently, we shall suppose that except for the initial elastic action, we are dealing with an ideal incompressible plastic material, with a zero elastic strain range. This is consistent of course with the surmise of the initial very high speed elastic stress wave and the "freezing" of the material when it unloads as the bending wave passes by. Furthermore, we shall suppose that there are no strain rate effects affecting the plastic flow. We shall suppose, therefore, that the data obtained from a tensile test of the diaphragm material in which the tensile stress is exhibited as a definite function of the natural elongational strain, are, in combination with general laws of plastic flow, sufficient to specify the plastic properties of the material.

Naturally the mathematical statement of the problem depends to a certain extent upon the choice of the type of plasticity theory, although in any case an apparently consistent formulation may result. As a matter of fact, for the type of motion herein investigated, no differences arise whether one applies the plastic deformation type of theory, or the plastic flow theory of plasticity.‖ The point of view of the latter is taken in this article.

With the foregoing picture in mind, it is found possible to develop a consistent mathematical formulation of the theory of the main motion so as to satisfy the fundamental laws of mechanics.¶ This is done in the following sections.

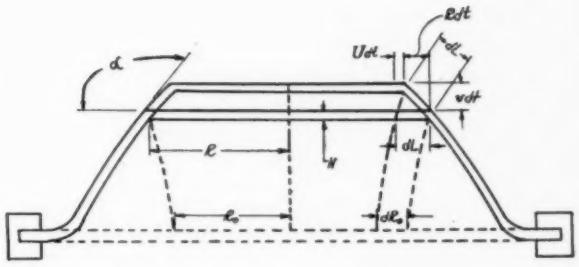


FIG. 3. A sketch of the diaphragm configurations at two successive instants, t and $t+dt$. The various quantities which enter into the several geometric and kinematic relations are portrayed here.

MATHEMATICAL FORMULATION OF THE THEORY

The basic suppositions and assumptions introduced in the preceding section lead to certain quantitative relations between the parameters which describe the configurations and state of the diaphragm material. The development of these relations can be divided into four main parts. Part 1 has to do with conservation of momentum, energy, and mass (volume) in the neighborhood of the bending wave. In Part 2, the equations of motion of the material in the central flat plastic portion are derived, and in Part 3, plasticity theory is introduced and applied. Part 4 is concerned with the specification of the initial state of the motion.

1. Conditions in the Neighborhood of the Bending Wave

Suppose that, at any time t, after the start of the motion, the distance of the bending wave from the center of the diaphragm is R, as shown in Fig. 3. Let dL be the width of an elementary ring of thickness H just ahead of the bending wave. During the time interval dt, this ring is swept over and, in effect, tilted by the bending wave as it is propagated inward. The radial distance traveled by the wave in this time is $-\dot{R}dt$, the negative sign being affixed because $\dot{R}$, the velocity of the bending wave, is negative. Let the material at a distance dL ahead of the bending wave have a radial velocity U outward, so that it moves a distance Udt in time dt. Hence, as is evident from Fig. 3, the rate at which material is swept over by the wave is

$$dL/dt=-\dot{R}+U. \tag{1}$$

As the wave travels by, the material ring of width dL in effect is tilted impulsively into a truncated conical ring of width dL and thickness H. In deforming, the inner edge of the annulus undergoes a displacement vdt normal to the original plane of the diaphragm while the

§ Henceforth, these latter are denoted as "plastic stresses" to distinguish them from the stresses arising from the constraints.

‖ For a discussion of these basically different theories see W. Prager, J. Appl. Phys. **19**, 540 (1948).

¶ The treatment of the initial elastic action is perhaps not quite as satisfactory, except when the hypothesis of a zero elastic strain range is rigidly maintained. This appears to be a subject for further investigation.

outer edge remains fixed. Actually, then, the bending wave travels a distance dL along the generator of the conical element while it is traveling inward along the radius a distance $-\dot{R}dt$. Hence, the rate at which the wave travels along the profile of the deformed diaphragm is given by

$$\frac{dL}{dt}=(\dot{R}^2+v^2)^{\frac{1}{2}}. \tag{2}$$

It is clear, from this and Fig. 3, that the generator of the conical element makes an angle α with the outward pointing radius, defined by the relations

$$dL/dt\cos\alpha=\dot{R}, \tag{3}$$

$$dL/dt\sin\alpha=v, \tag{4}$$

which together are equivalent to (2).

Let S_{RR}' be the total normal stress component parallel to a generator in the diaphragm at a point just behind the bending wave, and let S_{RH}' be the total shear stress component parallel to a plane element normal to the generator at this point as in Fig. 4. By total stress is meant here the sum of the plastic and the constraint stresses. These components exert forces of magnitude $S_{RR}'HRd\theta$ and $S_{RH}'HRd\theta$ respectively on the outer edge of a segment of width dL and thickness H which subtends an angle $d\theta$ at the center of the diaphragm.

In time dt, this segment loses all its momentum component, $\rho HRd\theta dLv$, normal to the original plane of the diaphragm. If the change in momentum is equated to the impulse $[-S_{RR}'HRd\theta\sin\alpha+S_{RH}'HRd\theta\cos\alpha]dt$ delivered in this direction in time dt, there results the relation

$$\rho(dL/dt)v=S_{RR}'\sin\alpha-S_{RH}'\cos\alpha. \tag{5}$$

Similarly, if S_R is the total principal radial stress component in the diaphragm material just ahead of the bending wave, as in Fig. 4, the net radial force component on the same segment of the annular element while it is being tilted is $-S_{RR}'HRd\theta\cos\alpha-S_RHRd\theta-S_{RH}'HRd\theta\sin\alpha$. In time dt this effects the removal of the radial momentum $\rho HRd\theta dLU$ from the segment. Again if the change in the radial momentum component is equated to the impulse, we obtain the relation

$$\rho(dL/dt)U=S_R+S_{RR}'\cos\alpha+S_{RH}'\sin\alpha. \tag{6}$$

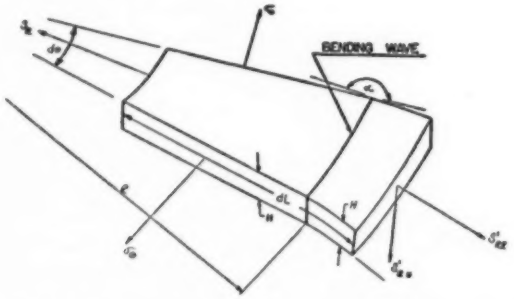


FIG. 4. Plastic and constraint stresses at the bending wave. The bending wave has traveled partly through a segment of an annular element, and the various stresses which serve to remove the momentum from this segment are shown.

As the bending wave sweeps inward over the element depicted in Fig. 4, the stress S_R does work of amount $-S_RHRd\theta Udt$ on it, while the stresses S_{RR}' and S_{RH}' being of the nature of constraint forces in the rigid stationary material behind the wave, do no work. The work which is done in the time dt equals the increase in kinetic energy, so that

$$\tfrac{1}{2}\rho\frac{dL}{dt}(U^2+v^2)=S_RU. \tag{7}$$

Several consequences of interest can be deduced quickly from Eqs. (1) through (7) inclusive. Recognizing from (1) and (2) that

$$U^2+v^2=2U(U-\dot{R})$$

and incorporating this in (7), and combining the result with (6), we find that

$$\rho\frac{dL}{dt}\dot{R}=S_{RR}'\cos\alpha+S_{RH}'\sin\alpha. \tag{8}$$

Substituting for $\dot{R}$ in (8) and for v in (5) from (3) and (4), we find, since the determinant

$$\begin{vmatrix}\sin\alpha, & -\cos\alpha\\ \cos\alpha, & \sin\alpha\end{vmatrix}\neq 0$$

that the speed of the bending wave with respect to the material on either side of it is

$$\frac{dL}{dt}=\left(\frac{S_{RR}'}{\rho}\right)^{\frac{1}{2}} \tag{9}$$

and that the shear stress behind the bending wave is

$$S_{RH}'=0. \tag{10}$$

Substituting the value of U obtained from (1) into (6) and utilizing (8) and (9), we find that

$$S_{RR}'=S_R \tag{11}$$

so that there is no stress discontinuity and no shock in this sense at the bending wave. In addition (3) and (4) may be combined to give

$$\tan\alpha=v/\dot{R}. \tag{12}$$

Similarly (1) and (2) yield

$$U^2-2U\dot{R}-v^2=0 \tag{13}$$

while utilization of $U^2+v^2=2U(U-\dot{R})$, obtained previously from (1) and (2), in (7), and then substituting for $(U-\dot{R})$ from (1) and for (dL/dt) from (2), yields

$$\rho(\dot{R}^2+v^2)=S_R. \tag{14}$$

Equations (9) through (14) inclusive are entirely equivalent to the original six independent relations

described by (1) through (7), inclusive. Clearly we may regard these equations as defining the six quantities dL/dt, S_{RH}', S_{RR}', α, $\dot{R}$, and U in terms of v and ρ, and S_R.

2. Equations of Motion in the Flat Central Region

The component velocity of any particle in the flat central region, perpendicular to the original diaphragm plane, is supposed to have the constant value v until the bending wave sweeps inward over it. This normal motion is quite independent of the radial motion, which in general might conceivably be much more complicated. The radial motion is of importance, for upon it depends, in part, the distribution of thinning in the diaphragm. Let us now consider this radial motion.

As mentioned previously, in order to avoid the mathematical complication of nonuniform thinning in the flat central region and the necessity of attempting to solve nonlinear partial differential equations of motion, we introduce a system of constraints which serve to maintain a uniform thickness H throughout this region. Consequently, at time $t+dt$, a disk which was of radius $R-dL$, and uniform thickness H at time t, has stretched into one of radius $R-dL+Udt$ and uniform thickness $H+\dot{H}dt$. Since its volume is conserved we are led to the equation

$$\dot{H}/H+2U/R=0. \tag{15}$$

A similar consideration of an interior disk of momentary radius $r<R$, at time t, and initial radius r_0, shows that

$$r/r_0=\partial r/\partial r_0=(h/H)^{\frac{1}{2}} \tag{16}$$

so that r/r_0 is a function of time only. Equations (16) will be termed the "constraint equations" for the flat central region.

Now the forces exerted on each material particle in the central region are due to the constraints and to the plastic stresses. We are supposing that the only normal plastic stresses in the central region are radial and circumferential stress components σ_r and σ_θ, respectively. On an elementary segment, of width dr and thickness H, which is a distance r from the center of the diaphragm where it subtends an angle $d\theta$, these stresses exert a net force

$$[\partial/\partial r(r\sigma_r)-\sigma_\theta]Hd\theta dr.$$

But from the constraint expressed by (16), it is clear that the radial and tangential strains, and hence the radial and tangential strain velocities are equal and functions of the time only. It follows from plasticity theory that the radial and tangential plastic stress components are equal and are functions of the time only (see Part 3). Hence, the plastic stresses exert no net accelerating force on the particles in the flat central region interior to the bending wave. The only possible accelerating forces in this region are the constraint forces; that is, we may define the outward accelerating force on the element of mass $\rho Hrdrd\theta$ by a differential quantity $d\Gamma d\theta\equiv(\partial\Gamma/\partial r)drd\theta$, except for elements just ahead of the bending wave. Combining (16) with the equation of motion of this element, we find that

$$\rho H^{\frac{1}{2}}r^2\frac{d^2}{dt^2}\left(\frac{1}{H^{\frac{1}{2}}}\right)=\frac{\partial\Gamma}{\partial r}. \tag{17}$$

At the bending wave itself we must have

$$RHS_R=RH\sigma_r+\Gamma_R \tag{18}$$

where Γ_R is the value of Γ when $r=R$, and is to be so chosen that the constraint forces do no work. This means that the rate of working by the whole set of constraint forces is zero, *viz.*,

$$\int_0^{2\pi}\int_0^R\frac{\partial r}{\partial t}d\Gamma d\theta-2\pi\Gamma_R U=0;$$

if this integration is carried out, using (15), (16), and (17), we see that

$$\Gamma_R=\rho H^{\frac{1}{2}}\frac{R^3}{4}\frac{d^2}{dt^2}\left(\frac{1}{H^{\frac{1}{2}}}\right). \tag{19}$$

Equation (17) may now be integrated to yield

$$\Gamma=\tfrac{1}{12}\rho H^{\frac{1}{2}}(4r^3-R^3)\frac{d^2}{dt^2}\left(\frac{1}{H^{\frac{1}{2}}}\right) \tag{20}$$

as the quantity defining the distribution of constraint forces in the flat central region.

It is interesting to note (and a check on the work) that if one now equates the total rate of decrease of kinetic energy for the whole diaphragm to the rate at which energy is being absorbed in plastic working of the material, *viz.* $-\sigma_r\dot{H}\pi R^2$, the resulting relation, after a bit of manipulation, is identical with (7), as it should be.

3. The Application of Plasticity Theory

To complete the discussion in the previous section it is necessary now to introduce certain stress and strain relationships in the flat central region. Because the rates of change of the principal strains are proportional to the principal strains themselves, as a result of the constraint relation (16), it is not important to formulate the plastic flow relationships in a completely general fashion.** It suffices to write these laws in terms of the principal stresses σ_1, σ_2, σ_3 and principal natural strains ϵ_1, ϵ_2, ϵ_3 in the form

$$\epsilon_1+\epsilon_2+\epsilon_3=0$$

and

$$\frac{\sigma_1-\sigma_3}{\epsilon_1-\epsilon_3}=\frac{\sigma_2-\sigma_3}{\epsilon_2-\epsilon_3}=\frac{\tau(\gamma)}{\gamma}.$$

** For such formulations, see, for example: Handelman, Lin, and Prager, Quar. Appl. Math 4, 397–407 (1947). In the present report, the plastic flow laws reduce to a form identical with that of the deformation theory of plasticity.

The first states the law of conservation of volume, while the second, as a consequence of the first, states that the principal strains (more generally the strain rates) are proportional to the deviations of the stresses from an isotropic stress state. The factor of proportionality, τ/γ, is the ratio of the octahedral shear stress

$$\tau=\tfrac{1}{3}[(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2]^{\frac{1}{2}}$$

to the octahedral shear strain

$$\gamma=\tfrac{1}{3}(\epsilon_1^2+\epsilon_2^2+\epsilon_3^2)^{\frac{1}{2}}$$

and, when the material is yielding plastically, the ratio is a function of γ, having a form determined empirically from a tensile test for example. According to our previous supposition, in a tensile test, the tensile stress σ_T, is related to the natural longitudinal strain, ϵ_L, by

$$\sigma_T=\sigma(\epsilon_L).$$

In a tensile test $\tau=(\sqrt{2}/3)\sigma_T$, while $\gamma=(\sqrt{2}/2)\epsilon_L$. Hence, we may rewrite the above relation as

$$\tau(\gamma)=(\sqrt{2}/3)\sigma(\sqrt{2}\gamma)$$

which is now applicable to a more general stress state than that occurring in a tensile test.

Let us now apply these laws to the central flat region of the diaphragm. In this case

$$\sigma_1=\sigma_r, \qquad \sigma_2=\sigma_\theta, \qquad \sigma_3=\sigma_H=0,$$

$$\epsilon_1=\epsilon_r=\log\frac{\partial r}{\partial r_0}, \qquad \epsilon_2=\epsilon_\theta=\log(r/r_0), \qquad \epsilon_3=\epsilon_H=\log\frac{H}{h},$$

where σ_H is the principal stress in the thickness direction, and ϵ_r, ϵ_θ, and ϵ_H are the natural (logarithmic) strains in the directions indicated by the subscripts. As a consequence of (16), $\epsilon_1=\epsilon_2=-\frac{1}{2}\epsilon_3$, so we find that

$$\sigma_r=\sigma_\theta \tag{21}$$

and

$$\sigma_r=\sigma(\log h/H). \tag{22}$$

Equality (21) has already been used in the previous section in obtaining (17).

4. The Specification of the Initial State

As the initial elastic stress wave front sweeps inward from the edge it may accelerate the material particles, leaving any real diaphragm in a state of plastic flow with the material flowing radially outwards. However, for the ideal material with which we have been dealing, with a zero elastic strain range, the elastic stress wave would travel at an infinite speed from the periphery to the center. Consequently, although the discontinuity in stress across the front of such a wave is quite high, jumping from zero to the yield stress $\sigma=\sigma(0)$, the actual impulse delivered to any material particle swept over by the front would be zero. This seems to be possibly a rather extreme case, however, and since the theory does permit the assumption of an initial radial velocity distribution of a certain kind, a rough estimate of the possible magnitude of this effect will be made as follows.

In accordance with the constraints placed on the diaphragm material, we shall assume that the motion, generated in the elastic deformation, can be approximated by one of a normal type with a linear radial velocity distribution which is zero at the center and greatest at the periphery. Then the radial coordinate r of a particle, initially at r_0, is

$$r=r_0\left(1+\frac{q(t)}{a}\right),$$

where q is the normal coordinate of the constrained motion, and "a" is the radius of the diaphragm. The elastic stresses for such a motion are

$$\sigma_r=\sigma_\theta=(E/1-\nu)q/a,$$

where E is Young's modulus and ν is Poisson's ratio. The motion begins at a time $t=0$, when the stress at the edge rises suddenly to the yield stress σ. Hence the equation of motion is, after calculating the equivalent mass from the kinetic energy,

$$\frac{\rho a}{4}\ddot{q}+(E/1-\nu)\frac{q}{a}=\sigma$$

to be solved subject to $\dot{q}(0)=0$, $q(0)=0$. We find that

$$\dot{q}(t)=2\sigma\left(\frac{1-\nu}{\rho E}\right)^{\frac{1}{2}}\sin\frac{2}{a}\left(\frac{E}{\rho(1-\nu)}\right)^{\frac{1}{2}}t,$$

$$q(t)=\frac{a\sigma(1-\nu)}{E}\left[1-\cos\frac{2}{a}\left(\frac{E}{\rho(1-\nu)}\right)^{\frac{1}{2}}t\right].$$

When $\dot{q}$ is a maximum, the stresses in the diaphragm reach the yield point, and from that time on, plastic flow takes over. This happens at a time

$$t_E=\frac{\pi a}{4}\left(\frac{\rho(1-\nu)}{E}\right)^{\frac{1}{2}};$$

the velocity distribution may be written

$$\frac{\partial r}{\partial t}=\frac{r_0}{a}\dot{q}(t_E)=v\frac{r_0}{a}\left(\frac{4\sigma}{\rho v^2}\cdot\frac{\sigma(1-\nu)}{E}\right)^{\frac{1}{2}}. \tag{23}$$

The time for complete plastic deformation of a diaphragm may be estimated roughly, considering the bending wave to travel along a radius with a speed

$$\dot{R}\cong-(\sigma/\rho)^{\frac{1}{2}}.$$

It reaches the center at a time

$$t_S\cong a(\rho/\sigma)^{\frac{1}{2}};$$

the ratio of t_E to t_S is thus

$$\frac{t_E}{t_S}\cong\frac{\pi}{4}\left(\frac{\sigma(1-\nu)}{E}\right)^{\frac{1}{2}}\cong 3\text{ percent}$$

for some steels.†† Similarly, elastic considerations give the radial velocity at the periphery,

$$\left.\frac{\partial r}{\partial t}\right]_{\text{Elastic}}=v\left(\frac{(1-\nu)\sigma}{E}\right)^{\frac{1}{2}}\left(\frac{4\sigma}{\rho v^2}\right)^{\frac{1}{2}}\cong 0.675\,v$$

while a rough order of magnitude of a typical plastic radial flow rate U may be obtained from (13):

$$U\cong\frac{v^2}{-2\dot{R}}\cong v\left(\frac{\rho v^2}{4\sigma}\right)^{\frac{1}{2}}\cong 0.055\,v,$$

so that

$$U\Big/\left.\frac{\partial r}{\partial t}\right]_{\text{Elastic}}\cong\frac{55}{675}\cong 8\text{ percent}.$$

From these results we see that although the initial elastic effects occur very rapidly, they may result in a radial velocity distribution of significantly large order of magnitude. Because of this, it is a bit unfortunate that the estimate (23) is not based on a firmer foundation; this appears to be a subject for further investigation.

The remaining initial conditions, other than (23) (which we now take to hold at $t=0$) may be listed. The initial radius of the bending wave is

$$R(0)=a; \tag{24}$$

the initial thickness is

$$H(0)=h; \tag{25}$$

the initial radial velocity at the bending wave is

$$U(0)=v\left(\frac{4\sigma}{\rho v^2}\cdot\frac{\sigma(1-\nu)}{E}\right)^{\frac{1}{2}}, \tag{26}$$

where $E=\infty$, for a material of zero elastic strain range.

Finally in order to define the final shape of the diaphragm profile, we introduce the distance $Z(t)$ of the central flat region from the initial plane of the diaphragm. Clearly,

$$Z=vt \tag{27}$$

so that once R is found as a function of t, elimination of t between R and Z will yield $Z=Z(R)$, the equation of the diaphragm profile. This concludes the mathematical formulation of the theory.

†† Typical values of parameters for obtaining these quick estimates are:

$\sigma=6\times10^4$ lb/in.^{2} $\quad\rho=0.75\times10^{-3}$ lb sec^2/in.^{4} $\quad h=0.05$ in.
$E=3\times10^7$ lb/in.^{2} $\quad v=10^3$ in./sec $\quad a=5$ in.
$\nu=0.3$

SUMMARY OF THE EQUATIONS OF MOTION

For convenience, the relevant defining relations are collected here:

$$(A)\quad\begin{cases}\dfrac{dL}{dt}=\left(\dfrac{S_R}{\rho}\right)^{\frac{1}{2}} & \Gamma_R=\rho H^{\frac{1}{2}}\dfrac{R^3}{4}\dfrac{d^2}{dt^2}\left(\dfrac{1}{H^{\frac{1}{2}}}\right)\\[2ex] \tan\alpha=\dfrac{v}{\dot{R}} & S_R=\sigma_r+\dfrac{\Gamma_R}{RH}\\[2ex] S_{RR}'=S_R & \sigma_r=\sigma_\theta=\sigma\left(\log\dfrac{h}{H}\right)\\[2ex] S_{RH}'=0 & Z=vt\end{cases}$$

$$(B)\quad\begin{cases}U^2-2U\dot{R}-v^2=0\\[2ex] \rho(\dot{R}^2+v^2)=\sigma\left(\log\dfrac{h}{H}\right)+\frac{1}{4}\rho R^2H^{\frac{1}{2}}\dfrac{d^2}{dt^2}\left(\dfrac{1}{H^{\frac{1}{2}}}\right)\\[2ex] \dfrac{\dot{H}}{H}+2\dfrac{U}{R}=0\end{cases}$$

$$(C)\quad\begin{cases}R(0)=a\\ H(0)=h\\ U(0)=v\left(\dfrac{4\sigma}{\rho v^2}\cdot\dfrac{\sigma(1-\nu)}{E}\right)^{\frac{1}{2}}.\end{cases}$$

In (A) the quantities on the left of the equations are defined in terms of ρ, h, v, $\sigma(\log h/H)$, t, R, and H. The relations (B) constitute a differential system of third order in H, U, and R, so that (C) furnishes the necessary three initial conditions. In order to solve (B), it is necessary, in addition, to have an explicit functional form for $\sigma(\log h/H)$. For example, if the ideal material is one which does not work harden, then we have simply

$$\sigma(\log h/H)=\sigma \tag{28}$$

where σ is a constant.

An Elementary Approximation to the Solution

In case the effect of the constraint forces, that is, acceleration of the material in the central flat region, is small throughout most of the motion, so that it may be neglected, the equations (B) take on a particularly simple form, especially if the material is one which does not work harden. Then the order of the set of differential relations becomes two, so that only two initial conditions, those on R and on H, can be applied. Although it is not necessary, it is also convenient to suppose that $\sigma/\rho v^2$ is a very large number compared to unity, which is certainly true for many cases of interest. For the typical values listed in a previous section $\sigma/\rho v^2=80$.

Let us denote the constant σ/ρ by c^2. Then, neglecting high powers of v/c, the solution of the equations may be

written

$$\left.\begin{aligned} U &= \frac{v^2}{2c} && \text{(a)}\\ R &= a - ct && \text{(b)}\\ H &= h\left(\frac{R}{a}\right)^{(v/c)^2} && \text{(c)}\\ Z &= \frac{v}{c}(a-R) && \text{(d)} \end{aligned}\right\}. \quad (29)$$

so that

These equations tell us that the radial speed U is a small constant at the bending wave, and that the speed of the bending wave from the edge to the center of the diaphragm is a constant, independent of the initially imposed normal velocity v. The thickness distribution in the deformed diaphragm given by Eq. (29)(c), shows a dimpling tendency at the center; in fact at the last moment the thickness becomes zero at the very center—no doubt a consequence of the idealizations and approximations. In any event, the variation with R is so rapid near $R=0$, the tiny pinhole would not actually be experimentally very apparent. Furthermore, this rapid variation of thickness indicates that at the last moment, when $R/a \ll 1$, the neglect of the constraint effects might be more serious than otherwise; however, this does not turn out to be the case in general, as shown in the next section. Equation (29)(d) shows that the diaphragm assumes a conical shape, as indicated by experiment, whose center deflection is proportional to the initial velocity v. This last is the basis for the use of such diaphragms as impulsive velocity indicators, as described earlier. Quantitatively there is also fair agreement between these results and experimental observations. Indeed, even the total time for deflection, which has been called the swing time t_S, of the diaphragm, and is given by

$$t_S = a/c,$$

is seen to be independent of v, to this order of approximation; computed values where higher powers of v/c are included agree rather well with the experimentally observed ones.

The results of this section on the elementary theory were implicit in the writer's first report on the diaphragm theory. In that report, as in this section, no serious attempt was made to estimate the effect of the motion in the central flat region on the explicit solution of the problem; that is, the nonlinear partial differential equations of motion for this region were not taken into account by the introduction of constraints.

The present article is a generalization of the former one, as will be seen in the next section; there the exact solution of the equations (B) and (C) are obtained and inform us as to the effect of radial motion and thinning.

THE EXACT SOLUTION OF THE EQUATIONS OF MOTION FOR A MATERIAL WITH NO WORK-HARDENING

Although the differential equations (B) are nonlinear, still they may be solved explicitly in a finite form involving a single quadrature, at least in the case for which there is no work-hardening. Let us introduce the nondimensional parameters

$$\begin{aligned} &\beta = R/a, \quad \zeta = vt/a = Z/a,\\ &\mu = U/v, \quad \delta = 4\sigma/\rho v^2,\\ &\eta = H/h, \quad \kappa = [(4\sigma/\rho v^2)\sigma(1-\nu)/E]^{\frac{1}{2}}, \end{aligned}$$

where, for the values suggested in a previous section $\delta = 320$, and $\kappa = 0$, or 0.675.

After a certain amount of algebra, equations (B), reduce to

$$\begin{aligned} \eta'/\eta &= -2(\mu/\beta) && \text{(a)}\\ 2\mu\beta' &= \mu^2 - 1 && \text{(b)}\\ (2-\mu^2)\beta' + \beta\mu\mu' &= (3-\delta)\mu && \text{(c)} \end{aligned} \quad (30)$$

where a prime means differentiation with respect to ζ. These equations are to be solved subject to the initial conditions (C) which may be rewritten:

$$\beta = \eta = 1, \quad \mu = \kappa, \quad \text{when} \quad \zeta = 0. \quad \text{(d)} \quad (30)$$

By dividing the sides of (30)(c) by the corresponding sides of (30)(b), the variable ζ is eliminated, and we are left with an integrable equation whose variables are separated, relating β, and μ. Similarly by multiplying the right side of (30)(a) by the left side of (30)(b) and conversely, and employing the previous integrable expression to eliminate β, we find an integrable relation involving η and μ. Finally ζ is found in quadrature form in terms of μ, and $\beta(\mu)$ from (30)(b). This explicit solution may be exhibited as

$$\begin{aligned} \frac{R}{a} = \beta(\mu) &= \left(\frac{\mu^2 - x_1}{\kappa^2 - x_1}\right)^{\frac{1}{2}(x_1-1)/(x_1-x_2)} \\ &\quad \times \left(\frac{\mu^2 - x_2}{\kappa^2 - x_2}\right)^{\frac{1}{2}(x_2-1)/(x_2-x_1)} && \text{(a)}\\ \frac{H}{h} = \eta(\mu) &= \left(\frac{\mu^2 - x_1}{\kappa^2 - x_1}\right)^{-2(x_1)/(x_1-x_2)} && (31)\\ &\quad \times \left(\frac{\mu^2 - x_2}{\kappa^2 - x_2}\right)^{-2(x_2)/(x_2-x_1)} && \text{(b)}\\ \frac{Z}{a} = \zeta(\mu) &= \int_{\kappa}^{\mu} \frac{2\mu^2}{(\mu^2 - x_1)(\mu^2 - x_2)}\beta(\mu)\,d\mu && \text{(c)} \end{aligned}$$

where x_1, x_2, are the smaller and larger (positive) roots respectively of the quadratic equation

$$x^2 - (2\delta - 3)x + 2 = 0 \quad (32)$$

and $\mu = U/v$. Unfortunately (31)(c) is not integrable in terms of elementary functions, in general; still a great

deal of information can be gained from an analysis of (30), (31), and (32), and from a consideration of special cases.

For positive values of δ, which are the only ones of physical interest, the roots of (32) occur in pairs of positive values for $\frac{3}{2}+\sqrt{2}\leqq\delta\leqq\infty$, and as pairs of negative values for $0\leqq\delta\leqq\frac{3}{2}-\sqrt{2}$ (a very small range). For intermediate values of δ, the roots are complex. We shall be interested only in values of $\delta\geqq\frac{3}{2}+\sqrt{2}$. At the lower limit of this range, corresponding to very large v, or small σ values, we have $x_1=x_2=\sqrt{2}$. As δ becomes very large, $\sqrt{2}>x_1\to 0$ and $\sqrt{2}<x_2\to\infty$.

Special Cases

A. A case of considerable interest arises when κ^2 coincides with one of the roots of (32). Now from (31) we see that the only possibility is for μ^2 itself to be constant and equal to κ^2. From (30)(b) we then find that either $0\leqq\mu=\kappa\leqq 1$ or $\mu=-\kappa\leqq -1$, in order for β' to be negative. The first case only is admissible, since the second precludes $\mu=\kappa$ initially. Thus we have

$$0\leqq\mu^2=\kappa^2=x_1=(\delta-\tfrac{3}{2})-((\delta-\tfrac{3}{2})^2-2)^{\frac{1}{2}}\leqq 1$$

and $\delta\geqq 3$.

The solution of (30) may now be written as

$$\left.\begin{aligned} R&=a-\frac{1-\kappa^2}{2\kappa}vt=a-\frac{1-\kappa^2}{2\kappa}Z && \text{(a)}\\ H&=h\left(\frac{R}{a}\right)^{4(\kappa^2)/(1-\kappa^2)} && \text{(b)}\\ U&=\kappa v && \text{(c)} \end{aligned}\right\} \quad (33)$$

and

$$\delta=4\frac{c^2}{v^2}=\frac{\kappa^4+3\kappa^2+2}{2\kappa^2}\geqq 3 \qquad \text{(d)}$$

where, it will be remembered, $c=(\sigma/\rho)^{\frac{1}{2}}$.

For $\kappa=1$, $\delta=3$, and $v=(2/\sqrt{3})c$, and the deformed diaphragm is cylindrical, i.e., it ruptures completely. (Possibly in practice, rupture would occur before this, due to excessive thinning at the center, or shearing at the edge; experiments indicated the central failure to occur first with properly designed edge restraints. The above value of v might be regarded as an upper limit.)

For other allowed values of $\kappa(=\mu)$, this solution is quite similar to the elementary approximation presented earlier, even quantitatively so, for large δ values. For then

$$\kappa\cong 1/\delta^{\frac{1}{2}}=v/2c\ll 1, \quad (\delta\gg 3)$$

and (33) reduces to (29).

B. Still another case, perhaps even the most important one for purposes of this paper, arises if we suppose that κ may have any desired positive value, within reason, and that δ is very large, so that x_1 is small and x_2 is large. We shall assume further that $\kappa^2/x_2\ll 1$, as is μ^2/x_2 (since μ^2 varies from κ^2 to x_1). Under these simplifying conditions it is possible to derive an approximation to the integral (31)(c) but in a somewhat indirect way, much too long to reproduce here in detail. We find that:

$$x_1\cong 1/\delta, \quad x_2\cong 2\delta\gg 1,$$

where δ, it will be remembered is $4c^2/v^2$. Then, provided δ is large enough

$$\left.\begin{aligned} \frac{Z}{a}&=\frac{vt}{a}=2Uv/(v^2-U^2)[1-(R/a)] && \text{(a)}\\ \frac{H}{h}&=\left(\frac{R}{a}\right)^{(v/c)^2} && \text{(b)}\\ \frac{U}{v}&=\frac{v}{2c}\left\{1+\left(4\kappa^2\frac{c^2}{v^2}-1\right)\left(\frac{R}{a}\right)^{16(c/v)^2}\right\}^{\frac{1}{2}} && \text{(c)} \end{aligned}\right\} \quad (34)$$

It can be shown that this approximation is uniform over the range of U/v from κ to $v/2c$, and the range of R/a from 1 to 0; but the nearer κ is to 1, the larger δ must be. Now we may distinguish several possibilities. If $\kappa=0$, which is the case of a material which has strictly no elastic strain range, we find that the shape of the deformed profile, described by (34)(a), is conical near the center, with a center deflection identical with that given by the elementary theory. At the edge, the slope is zero (as required by (30)(b)). The thickness distribution is identical with that given by the elementary approximation, while from (34)(c), the radial velocity increases very rapidly from zero to $v^2/2c$. As κ increases the solution approaches the elementary approximation, the slope at the edge increasing, until when $\kappa=v/2c$, the shape is exactly conical as in (29). As κ increases further, the slope at the edge becomes greater, the center deflection remaining the same, however, as does the thickness distribution. The radial velocity U, at the bending wave, decreases, again very rapidly, from κv to $v^2/2c$. These results all seem physically quite reasonable and might have been expected on such grounds. A calculation indicates that there is actually very little other effect of different κ values on the profile, as long as δ is very large.

Apparently, then, we may conclude that the elementary approximation is even more reliable than could have been hoped for, as long as there is no work hardening; indeed, under the conditions noted, the entire discussion concerning equations (29) is applicable. In the next section we consider some effects of work hardening.

CONSIDERATIONS OF WORK-HARDENING

In line with the results of the last section, it seems apropos, when introducing the complication of work hardening, to neglect the radial motion, insofar as its inertial effects are concerned. This was done in the development of the elementary approximation in the previous section, as it is in the following.

FIG. 5. Profiles of completely deformed diaphragms. Curves *a* denote diaphragms whose material does not work-harden while *b* refers to diaphragms whose material work-hardens. The subscripts 1 and 2 refer to different initial conditions in which the initial velocity v was taken as 10^3 in./sec. and 2×10^3 in./sec, respectively. The effect of work-hardening in decreasing the center deflection is shown, and in rounding off the apex of the conical shape.

Now the stress function of a material which work hardens may be written

$$\sigma\left(\log\frac{h}{H}\right)=\sigma+\omega\log\frac{h}{H} \tag{35}$$

where as before σ is the yield stress and ω is the work hardening stress per unit natural strain in a tensile test, and has a value of approximately 400×10^3 lb/in.^{2} for some medium steels. Such a stress-strain relation is not at all inconsistent with many empirical data.

After introducing (35) into the general equations (B), with the constraint forces deleted, the differential equations of motion can be written in terms of the previous non-dimensional notation with the addition of

$$\epsilon=\omega/(\rho v^2).$$

If, furthermore, we let

$$\xi=\left(\frac{\delta}{4}-1+\epsilon\log\frac{1}{\eta}\right)^{\frac{1}{2}}$$

the equations take the simple form

$$\mu^2+2\mu\xi-1=0 \tag{a}$$

$$\frac{d\beta}{\beta}=-\frac{1}{\epsilon}\frac{\xi^2}{\mu}d\xi \tag{b}$$

$$d\zeta=-d\beta/\xi \tag{c}$$

(36)

where for the time being ξ, not ζ, may be considered to be the independent variable. Although these equations can be solved explicitly with but one indicated quadrature remaining (from (36)(c)), the solutions are so complicated that it is more instructive to consider a special case in which the approximation $\xi\gg1$ may be made. It is clear that for the typical values we have been utilizing this is the case (e.g., $\delta/4\cong80$). Thus (36)(a) immediately yields

$$\mu=-\xi+(\xi^2+1)^{\frac{1}{2}}\cong\frac{1}{2\xi}. \tag{37}$$

After some manipulation, the solutions, to this order of approximation, can be rewritten with β, or R/a, as the independent variable. Thus, with $c^2=\sigma/\rho$, as before, and $d^2=\omega/\rho$, we have

$$U=\frac{v^2}{2c}\{1-\phi\}^{-\frac{1}{2}} \tag{a}$$

$$H=h\exp-\frac{c^2}{d^2}\{(1-\phi)^{\frac{1}{2}}-1\} \tag{b}$$

$$Z=vt=-\frac{v}{c}\int_a^R\{1-\phi\}^{-\frac{1}{2}}dR \tag{c}$$

(38)

where $\phi\equiv2(v^2d^2)/c^4\log R/a$. It is interesting to note that as $d\to0$, that is, the work-hardening stress becomes small, the above solution reverts to the elementary approximation (29), since ϕ vanishes like d^2 in the limit. One major difference stands out between the above and the previous solutions—at the center of the deformed diaphragm, (dZ/dR) vanishes. That is, the apex of the conical shape is rounded off, in complete qualitative agreement with the observations. It is also of interest to observe that the center deflection of the deformed diaphragm is somewhat less than that of a diaphragm whose material does not work-harden, but which has the same yield limit and initial conditions. These observations are illustrated by the diagrams of Fig. 5, in which are compared the deformed diaphragm profiles for two materials, one of which does not work-harden. The profiles of the diaphragms whose material work-hardens were calculated by integrating numerically (38)(c) (which, incidentally, may be put in the form of an incomplete Γ-function). It is probable, that by taking account of strain-rate effects, which have not yet been incorporated into plasticity theory, an even more marked rounding off would occur at the center than is evident from Fig. 5.

One should note that, as a result of neglecting the constraint forces in the approximate solution for a work-hardening material the initial value of U is completely determined, and equal to $v^2/2c$, as in the elementary theory. A more general solution including the constraint terms would of course allow this initial value to be given arbitrarily. Although such a solution has not been investigated, it is likely that U would quickly approach the value given by (38)(a), after performing its initial task of increasing or decreasing the slope of the diaphragm profile near the edge.

CONCLUSIONS

In partial conclusion, before summing up the accomplishments of this theory, it might be well to point out some of its shortcomings. First, the constraint forces imposed on the central flat portion are seen actually to yield an infinite constraint tension or stress just at the center of the diaphragm. However, other idealized theories have frequently led to similar paradoxes; e.g.,

the pressure (which is actually a constraint stress) imposed by the assumption of incompressibility is infinite at a sharp corner in the theory of potential flow of a liquid. Second, the imposition of rigidity on the material behind the bending wave is an obvious artificiality; finally, the assumption that the bending wave is an actual discontinuity in slope is not strictly in accord with the facts. It would indeed be pleasant to be able to relax some or all of these idealizations; however, it is clear that, having recognized the artificialities for what they are, it has been quite necessary to invoke them in this paper in the name of mathematical simplicity.

Equations (*A*), (*B*), (*C*) together with an empirically determined stress-strain relation [analogous to (28) or (35)], specify completely the motion and plastic deformation of any diaphragm of radius a, thickness h, and initial velocity v. For instance, one may calculate the following:

1. The radius R of the bending wave as a function of the time.
2. The diaphragm profile at each instant.
3. The thickness distribution.
4. Displacement—time curves of particles in the diaphragm.
5. Stress and strain distributions.
6. The center deflection as a function of v.
7. The total time for the deformation to take place.

The viewpoint presented here admittedly has led to a kind of short-cut procedure which is designed to circumvent certain mathematical difficulties inherent in a more rigorous theory. Because of this method of attack, it may be necessary in the future, in order to make the results more generally applicable, to re-examine some of the basic assumptions made herein. Outstanding among these is the question of the initial radial velocity and stress distribution (although the results seem to indicate that this has only a slight effect on the final answer). Other investigations might consider such effects as the magnitude of the energy absorbed at the bending wave, equilibrium of the region behind the bending wave, and the possibility (which seems small) of plastic flow in it, and the possible nonuniform radial flow and tension in the central flat region.

To conclude, the theory presented herein describes a dynamically possible mode of motion, and one which can be handled simply, in a mathematical sense. Because the theory is based, for the most part, on experimentally observed facts, there is a strong possibility that the motion derived from it approximates closely that of a diaphragm deforming plastically after having been given an initial uniform velocity normal to its original plane.

In acknowledgment, the writer expresses his appreciation to Professor E. H. Kennard for his many useful suggestions and helpful criticisms in regard to the original development of this theory; to Mrs. E. L. Miller who performed the calculations leading to Fig. 5; to Miss M. Pfeiffer in helping prepare the final manuscript; to Mr. M. Storm for proofreading and checking the mathematics.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Improvements in Visual Depiction of Magnetic Lines of Force by Means of a Gas Discharge

F. Blaha and J. A. Schedling
First Physics Institute, University of Vienna, Austria
(Received March 16, 1950)

The article shows how lines of force of various magnetic fields can be visually depicted in three dimensions by means of a gas discharge. The method is illustrated by several photographs proving especially its applicability when investigating fields of small geometrical dimensions.

INTRODUCTION

RECENTLY several papers have been published describing particular appearances of gas discharges in magnetic fields,[1] one of them dealing with results obtained when using parts of a magnet as cathode.[2] In the latter case the discharge appears as luminous bands, in the cathode region of the gas discharge, which follow the magnetic lines of force. This suggested the use of the phenomenon as a means of mapping magnetic fields.

The authors have worked out an improved method which enables one to make visible several chosen lines of force at once (i.e. the entire field need not be mapped with multiple-exposure photographs)[1] and to their full extent between the opposite surfaces of the pole pieces. Furthermore, the new method turned out to be of decided advantage in investigating complicated shaped fields especially those of geometrically small dimensions, the mapping of which is generally difficult.

APPARATUS

The set-up used is shown in cross section in Fig. 1. The containing vessel is large enough so that the magnet can be placed directly in the discharge area and also so that the effect of the walls on the discharge will be small.

[1] *Electronics, The Electron Art*, edited by Frank Rockett, December, 1948.
[2] F. Ehrenhaft, Phys. Rev. **75**, 1628 (1949). F. Ehrenhaft, Acta Physica Austriaca, Bd IV, 1, 1950.

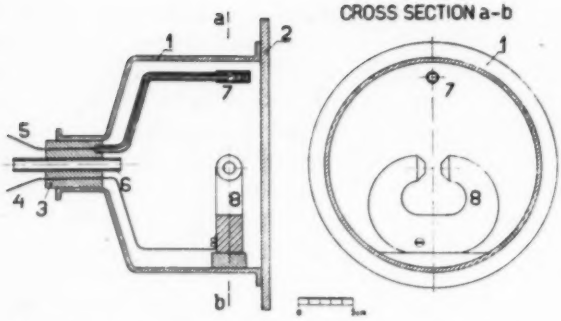


FIG. 1. Set-up used: 1. Bell jar; 2. Glass plate; 3. Stopper; 4. Lead to cathode; 5. Lead to anode; 6. Tube for evacuation; 7. Anode; 8. Magnet (cathode).

As cathode a permanent magnet was used; the anode, consisting of a plate or a cylinder of Cu, Al or some other metal, may be placed anywhere within the vessel.

The field strengths of the magnets tried amounted to 500–5000 gauss. The voltage between cathode and anode ranged from 400–800 volts, the current of discharge from 5 to 40 ma d.c.

The pressure in the vessel may be varied from 10 to 0.01 mm Hg, the gases used being air, carbon dioxide, etc. depending on the color desired.

PRELIMINARY INVESTIGATIONS

For the first part of the experiments a permanent magnet was equipped with two pole pieces shaped as truncated cones which were covered with a thin layer of Apiezon grease. It may be remarked that the magnet and the pole pieces are at the same electric potential, namely that of the cathode. The parts of the discharge in the region of the anode are normal, though the positive column is bent by the stray magnetic field present. In the space between and around the pole pieces appear intense luminous lines stretching from surface to surface, which apparently follow the magnetic lines of force (compare Fig. 2), and closer investigation showed their actual coincidence.

These lines often stay unchanged for minutes. The cross section of a single line increases with decreasing magnetic field strength. The light intensity seems to be weaker at the pole pieces.

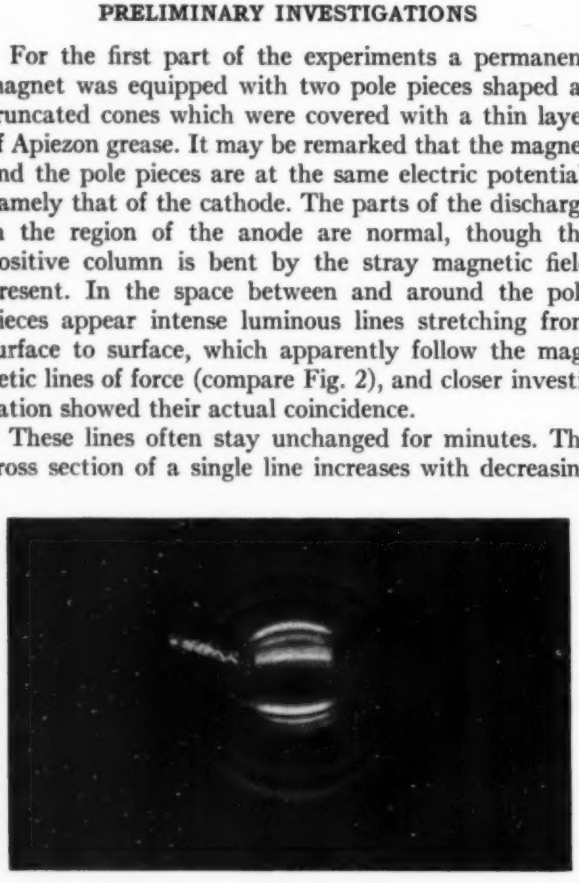


FIG. 2. Pole pieces shaped as truncated cones, covered with a thin layer of Apiezon grease.

FIG. 3. Pole pieces as in Fig. 2, only the upper halves being covered with grease.

At the moment when a new line appears, at its point of origin an intense spark of blue-green color can be seen for a short time. The points of origin of such lines cannot be fixed in advance, their random distribution seems to depend on the particular local conditions of the grease coating.

As to the function of the Apiezon grease it seems to be to screen off the normal glow discharge and to favor a sort of arc discharge at some points. This may be deduced from Fig. 3, where only the upper halves of the pole pieces were covered with Apiezon grease; here the lines are clearly seen. In the space between the lower bare halves of the pole pieces no single lines following the magnetic lines of force can be observed. Instead the space is filled by an intense general glow containing the layers typical of the cathodic parts of a discharge. In the course of these experiments the grease was first partly replaced by a perforated solid dielectric material as a cover. Figure 4 shows the space between the poles screened off solely by two sheets of mica bearing two holes, whereas the remaining curved surface is still covered with grease. Here the holes become the points of origin for the luminous bands. In this way one can fix in advance the number of lines and their starting points in any region of the pole surface the magnetic field of which is to be investigated.

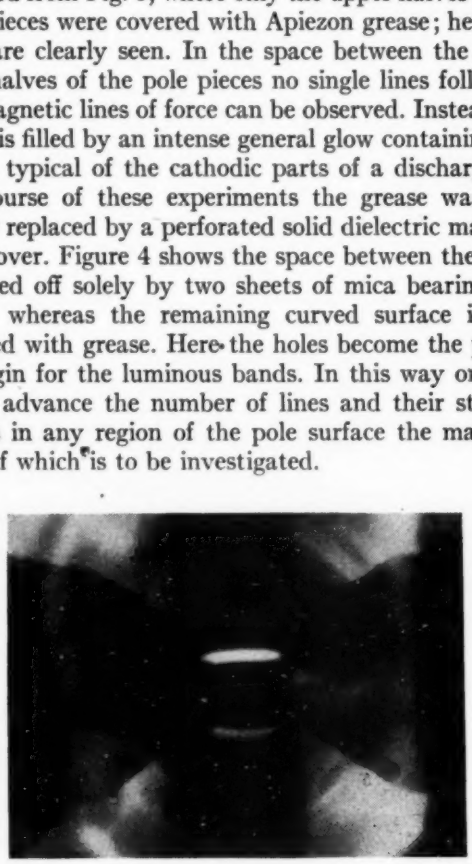


FIG. 4. Pole pieces partly screened off by sheets of mica.

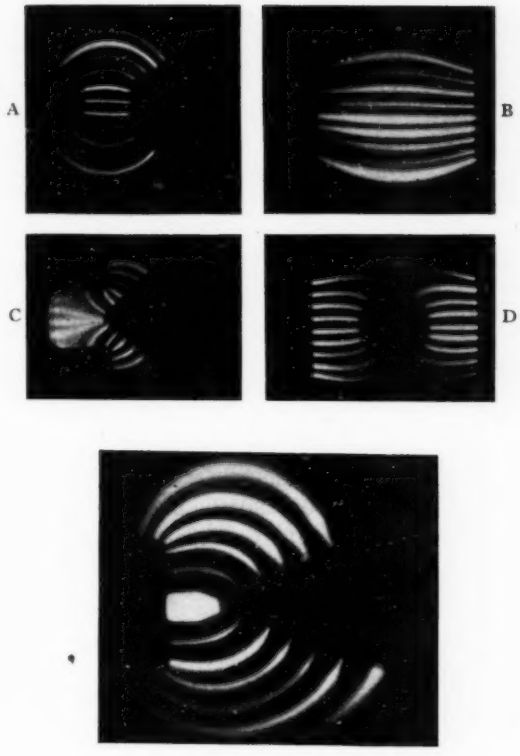


FIG. 5. Mapping of magnetic fields. Several types of pole pieces: A. Two truncated cones; B. Two planes; C. Wedge opposite a hole; D. Iron ring put into the field of B; E. Cone opposite a plane.

APPLICATIONS

In studying this method of screening, the entire surface of the various kinds of pole pieces used was covered with dielectric materials such as varnish, mica, plastics, etc., the covers having small holes of diameters of a fraction of a millimeter at points where the lines are desired.

Examples illustrating the type of information about the structure of the magnetic field which may be obtained from the method are shown in Fig. 5 (scale approximately 1:1). In all these examples the holes in the cover are adjusted in one plane normal to the direction of observation, for better photography. By suitable disposition of the holes a picture of the lines of force in all three dimensions may be obtained.

CONCLUDING REMARKS

Some hints on the formation of the relatively distinct luminous lines which are formed in more complicated

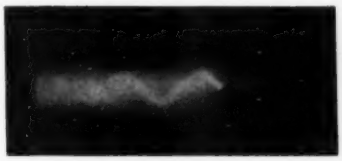


FIG. 6. Small cathode surface placed in the homogeneous magnetic field. The angle between the surface and the direction of the magnetic field being approximately 45°.

electric and magnetic fields may be deduced from Fig. 6. In this modification, not the magnet but the flat end of a thin copper wire which is sealed in a glass tube serves as cathode, the magnet itself being at arbitrary potential. The angle between the normal of this surface and the direction of the magnetic field is approximately 45°, the cathode being located in the homogeneous part of the field.

In Fig. 6 is seen the helix (a) (compare Fig. 7) the

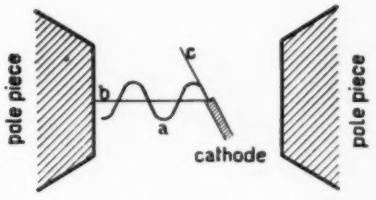


FIG. 7. Scheme of Fig. 6.

axis of which coincides with the magnetic line of force passing through the surface of the cathode. The diameter of the helix depends on magnetic field strength, the sense of the rotation indicating negatively charged particles of certain velocity. In addition, coaxial with the helix a weak band (b) can be visually observed (Fig. 6 does not show it because of the general illumination) which follows immediately the magnetic lines of force without having any visible helical structure.

Furthermore, there is visible a short straight ray (c) directed normal to the surface of the cathode. Figure 7 shows schematically the above described observations.

Varying the inclination of the cathode the helix changes diameter and pitch, and finally, when the normal of the cathode surface coincides with the magnetic field, the helix becomes a straight line parallel to the lines of force.

ACKNOWLEDGMENT

We wish to express our appreciation to Mr. W. Bergmann for his assistance in taking some pictures and carrying out the experiment concerning Fig. 6.

Electric Dipoles in the Presence of Elliptic and Circular Cylinders*

WINSTON S. LUCKE
Stanford Research Institute, Stanford, California
(Received May 5, 1950)

A theoretical examination is made of the problem of the electromagnetic fields of electric dipoles in the presence of an infinitely long, perfectly conducting cylinder. Application of the Green's function method to this problem yields expressions for the fields in terms of an integration in the complex plane. Simplified expressions for far zone fields are obtained through the known result of the complex integration for the far zone conditions. Results are given for both circular and elliptic cylinders. In the latter case, it is possible to treat the problem of dipoles in the presence of a ground plane which is finite in one dimension and infinite in the other, since this is equivalent to an elliptic cylinder with a vanishing minor axis. Polar diagrams of calculated far zone fields are included for a number of cases.

INTRODUCTION

THE problem of the electromagnetic fields of electric dipole radiators in the presence of an infinitely long, perfectly conducting cylinder of elliptic or circular cross section has attracted interest in the past which has led to a solution for the far zone field.[1] The analysis presented in this paper represents the application of the Green's function method to this problem, leading to expressions for the complete field in the form of an integral in the complex plane. Although the complex integration is formidable for the field at any point, it is shown that the imposing of far zone conditions leads to a simplification of the integral. By means of this simplification, expressions for the far zone field have been derived which are in agreement with previous results.[1]

Polar diagrams of the results of sample calculations are included. Among them is the case of a dipole in the presence of a flat strip of infinite length and finite width, i.e., a degenerate elliptic cylinder. This form of the problem of a finite ground plane gives results which agree qualitatively with experimental results for circular ground planes. Other sample calculations dealing with elliptic cylinders of mild eccentricities give results which may be applied to the study of the effect of wings, fuselage, and tail surfaces on the radiation properties of antennas mounted on aircraft or missiles.

GREEN'S FUNCTION METHOD OF SOLUTION

This analysis of the fields of electric dipoles of arbitrary orientation with respect to a perfectly conducting infinite cylinder (see Figs. 1 and 2 for examples of two orientations) is based upon the Green's function method. As discussed in detail by Morse and Feschbach[2] the application of this method to the solution of electromagnetic field problems leads to expressions for the field vectors within a region in terms of the currents and charges in the region and the fields on the surface of the region. Stratton[3] gives a general form of these

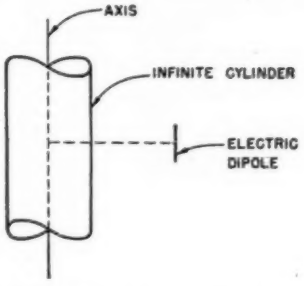


FIG. 1. Electric dipole parallel to the axis of an infinite cylinder.

expressions which may be applied to the present problem. Stratton shows that the electric and magnetic vectors, $\mathbf{E}(r)$ and $\mathbf{H}(r)$, within a region V bounded by

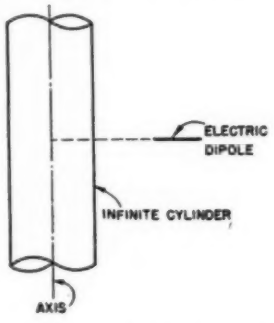


FIG. 2. Electric dipole in a plane perpendicular to the axis of an infinite cylinder.

a surface S (with normal n) are given by[4]

$$\mathbf{E}(r)=\int_V [i\omega\mu\mathbf{J}(r')G(r,r') + (1/\epsilon)\rho(r')\nabla' G(r,r')]dv' - \int_S \{i\omega\mu[n\times\mathbf{H}(r')]G(r,r') + [n\times\mathbf{E}(r')]\times\nabla' G(r,r') + [n\cdot\mathbf{E}(r')]\nabla' G(r,r')\}da', \quad (1)$$

* The research reported in this paper was made possible through the support extended the Stanford Research Institute by the U. S. Air Force, Contract No. AF 19(122)78.

[1] P. S. Carter, Proc. Inst. Radio Engrs. 31, 671–693 (1943). Carter treated the problem of electric dipoles in the presence of circular cylinders by a reciprocity method. It is understood that he has applied the same method to problems involving elliptic cylinders although the results have not been published.

[2] P. M. Morse and H. Feschbach, *Methods of Theoretical Physics* (Technology Press, Cambridge, Massachusetts, 1946).

[3] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1945), p. 466.

[4] m.k.s. units and a time dependence of $e^{-i\omega t}$ are assumed.

$$\mathbf{H}(r)=\int_V [\mathbf{J}(r')\times\nabla' G(r,r')]dv' + \int_S \{i\omega\epsilon[n\times\mathbf{E}(r')]G(r,r') - [n\times\mathbf{H}(r')]\times\nabla' G(r,r') - [n\cdot\mathbf{H}(r')]\nabla' G(r,r')\}da', \quad (2)$$

in which $\mathbf{J}(r)$ and $\rho(r)$ are the current density vector and the charge density scalar, respectively. The operator ∇' involves differentiation with respect to the primed coordinates. The Green's function, $G(r,r')$, satisfies the inhomogeneous wave equation

$$(\nabla^2+k^2)G(r,r')=-\delta(r-r'), \quad (3)$$

in which $\delta(r-r')$ is the three-dimensional Dirac delta-function. The solutions of Eq. (3) are arbitrary to the degree that the form of $G(r,r')$ can be chosen to satisfy prescribed boundary conditions, such as $G(r,r')=0$ or $\partial G(r,r')/\partial n=0$ on an arbitrary surface.

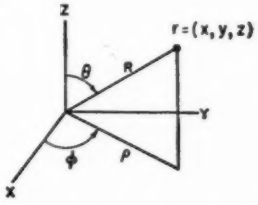


FIG. 3. Coordinate systems.

When these results are applied to the present problem, the surface S of Eqs. (1) and (2) is assumed to consist of an infinite, perfectly conducting cylindrical surface and the sphere at infinity. A cylindrical surface of circular cross section, in circular cylinder coordinates, ρ, ϕ, and z (see Fig. 3), is given by $\rho=a$, in which a is a constant. In elliptic cylinder coordinates, which are determined by the relations[5] (see Fig. 4),

$$x=\tfrac{1}{2}d\cos\phi\cosh\psi,$$
$$y=\tfrac{1}{2}d\sin\phi\sinh\psi,$$
$$z=z,$$
$$d=\text{distance between foci.}$$

To simplify the notation, the defining equations,

$$u=\cosh\psi, \quad v=\cos\phi, \quad (4)$$

are introduced. In these coordinates, therefore, a cylindrical surface of elliptic cross section is determined by $u=\cosh\psi=a$ in which a is a constant. In spherical coordinates, R, θ, and ϕ, the sphere at infinity is designated by $R\to\infty$. Electric dipoles, represented by the current density vector, $\mathbf{J}(r)$, and the charge density, $\rho(r)$, are assumed to be located within the volume V.

In order to simplify the expression for $\mathbf{E}(r)$ in Eq. (1), advantage is taken of the arbitrary nature of the expressions for the Green's function by setting $G(r,r')$ equal to zero on the cylindrical surface. By introducing this

[5] The notation of J. A. Stratton, P. M. Morse, L. J. Chu, and R. Hutner, *Elliptic Cylinder and Spheroidal Wave Functions* (John Wiley and Sons, Inc., New York, 1941), p. 53, hereafter designated by SMCH, has been adopted.

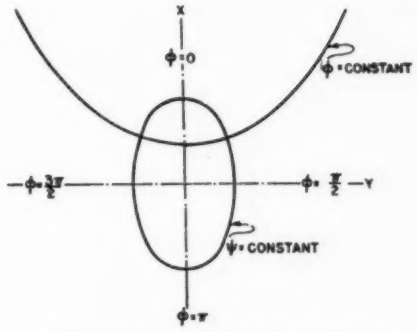


FIG. 4. Elliptic cylinder coordinates.

form of the Green's function in Eq. (1), assuming that the tangential electric field on the cylindrical surface vanishes, and noting that the integration over the sphere at infinity vanishes (see Appendix), it is found that Eq. (1) yields the z component of the electric vector which is given by

$$E_z(r)=\int_V [i\omega\mu J_z(r')G_1(r,r') + (1/\epsilon)\rho(r')(\partial/\partial z')G_1(r,r')]dv', \quad (5)$$

in which G_1 is that form of the Green's function which vanishes at ρ' or $u'=a$.

Similarly, the z component of the magnetic vector, determined by Eq. (2) is found to be

$$H_z(r)=\int_V [\mathbf{J}(r')\times\nabla' G_2(r,r')]_z dv', \quad (6)$$

in which G_2 is that form of the Green's function which satisfies

$$\partial G_2/\partial\rho'=0, \quad \rho'=a, \quad \text{or} \quad \partial G_2/\partial u'=0, \quad u'=a. \quad (7)$$

EVALUATION OF GREEN'S FUNCTIONS

In order to evaluate the integrals of Eqs. (5) and (6), it is necessary to find the appropriate forms of the Green's Function. These forms are obtained from the Morse and Feschbach[2] solutions, which satisfy Eq. (3), when modified to meet the prescribed boundary conditions. They are:

Circular cylinder,

$$G_1(r,r')=(i/8\pi)\int_C \sum_{m=-\infty}^{\infty} e^{im(\phi-\phi')}e^{i\alpha(z-z')}\frac{H_m^{(1)}(\beta\rho)}{H_m^{(1)}(\beta a)} \times[J_m(\beta\rho')H_m^{(1)}(\beta a) - H_m^{(1)}(\beta\rho')J_m(\beta a)]d\alpha, \quad (8)$$

$$G_2(r,r')=(i/8\pi)\int_C \sum_{m=-\infty}^{\infty} e^{im(\phi-\phi')} \times e^{i\alpha(z-z')}\frac{H_m^{(1)}(\beta\rho)}{(\partial/\partial\rho)[H_m^{(1)}(\beta\rho)]_{\rho=a}} \times\{J_m(\beta\rho')(\partial/\partial\rho)[H_m^{(1)}(\beta\rho)]_{\rho=a} - H_m^{(1)}(\beta\rho')(\partial/\partial\rho)[J_m(\beta\rho)]_{\rho=a}\}d\alpha. \quad (9)$$

Elliptic cylinder,

$$G_1(r,r')=(i/2\pi)\int_C\sum_{l=0}^{\infty}\frac{Se,\,o_l(\beta d/2,v)Se,\,o_l(\beta d/2,v')}{Ne,\,o_l}\frac{He,\,o_l^{(1)}(\beta d/2,u)}{He,\,o_l^{(1)}(\beta d/2,a)}\times\begin{Bmatrix}He,\,o_l^{(1)}(\beta d/2,a)Je,\,o_l(\beta d/2,u')\\-Je,\,o_l(\beta d/2,a)He,\,o_l^{(1)}(\beta d/2,u')\end{Bmatrix}e^{i\alpha(z-z')}d\alpha, \quad (10)$$

$$G_2(r,r')=(i/2\pi)\int_C\sum_{l=0}^{\infty}\frac{Se,\,o_l(\beta d/2,v)Se,\,o_l(\beta d/2,v')}{Ne,\,o_l}e^{i\alpha(z-z')}\frac{He,\,o_l^{(1)}(\beta d/2,u)}{(\partial/\partial u)[He,\,o_l^{(1)}(\beta d/2,u)]_{u=a}}\times\begin{Bmatrix}Je,\,o_l(\beta d/2,u')(\partial/\partial u)[He,\,o_l^{(1)}(\beta d/2,u)]_{u=a}\\-He,\,o_l^{(1)}(\beta d/2,u')(\partial/\partial u)[Je,\,o_l(\beta d/2,u)]_{u=a}\end{Bmatrix}d\alpha, \quad (11)$$

in which $\beta=[k^2-\alpha^2]^{\frac{1}{2}}$, $k=2\pi/\lambda$, and C is the contour[6] in the complex plane of α shown in Fig. 5. In the circular case J_m and $H_m^{(1)}$ are Bessel functions. In the elliptic case the notation e, o denotes summation over the even (e) and odd (o) terms such that there are two terms for each value of l in the summations of Eqs. (10) and (11). The functions $Je,\,o_l$, $He,\,o_l$, and $Se,\,o_l$ represent either Je_l or Jo_l, $He_l^{(1)}=Je_l+iNe_l$ or $Ho_l^{(1)}=Jo_l+iNo_l$, Se_l or So_l, respectively. The latter terms are solutions of the wave equation in elliptic cylinder coordinates. They are defined and tabulated by Morse and Feschbach[2] and SMCH.[5] The term $Ne,\,o_l$, is a normalization constant which is denoted by N in the cited references. Examination of Eqs. (8)–(11) shows that all forms of the Green's function satisfy the prescribed boundary conditions. Furthermore, it may be shown that all forms satisfy Eq. (3) by the methods demonstrated by Morse and Feschbach.[2]

Insertion of the appropriate forms of G_1 and G_2 in Eqs. (5) and (6) determines the electromagnetic field at the point r caused by any current distribution, $\mathbf{J}(r')$, in the presence of an infinite cylinder. The Green's function, however, contains an integral in the complex plane whose evaluation is laborious for arbitrary points r and r'. To avoid this complexity the far zone ($R\gg R'$) condition is chosen. For this case, it has been shown that the integral in the complex plane yields a simple result.

To determine the far zone fields, the asymptotic forms of the appropriate Green's function are substituted in Eqs. (5) and (6). These forms are dependent upon the following asymptotic expressions:

$$\begin{aligned}&H_m^{(1)}(\beta\rho)\sim(2/\pi\beta\rho)^{\frac{1}{2}}e^{i[\beta\rho-(2m+1)\pi/4]} && \rho\to\infty,\\ &He,\,o_l^{(1)}(\beta d/2,u)\sim(2/\beta du)^{\frac{1}{2}}e^{i[(\beta du/2)-(2l+1)\pi/4]} && u\to\infty.\end{aligned} \quad (12)$$

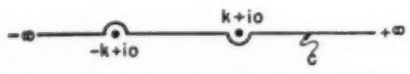


FIG. 5. Path of integration in the complex plane of α.

[6] A discussion of the contour C is contained in C. Papas, J. Appl. Phys. **20**, 437–440 (1949).

Since $du/2\sim\rho$ for large u, it follows that

$$He,\,o_l^{(1)}(\beta d/2,u)\sim(1/\beta\rho)^{\frac{1}{2}}e^{i[\beta\rho-(2l+1)\pi/4]}. \quad (13)$$

When Eqs. (12) and (13) are substituted in Eqs. (8)–(11), it is evident that the various forms of the Green's function depend on integrals of the form

$$\int_C\frac{\exp i[(k^2-\alpha^2)^{\frac{1}{2}}\rho+\alpha(z-z')-(2m+1)\pi/4]}{([k^2-\alpha^2]^{\frac{1}{2}}\rho)^{\frac{1}{2}}f([k^2-\alpha^2]^{\frac{1}{2}})}d\alpha,$$

in which $f([k^2-\alpha^2]^{\frac{1}{2}})$ is a function of ϕ, ϕ', ρ and $[k^2-\alpha^2]^{\frac{1}{2}}$. This integral has been evaluated by the method of steepest descents and the result which is given in a different form by Papas,[6] in the present notation, is

$$\sim(2\pi)^{\frac{1}{2}}e^{i(m+1)\pi/2}\frac{e^{ikR}}{R}\frac{e^{-ikz'\cos\theta}}{f(k\sin\theta)}. \quad (14)$$

When use is made of this result the forms of the Green's function in the far zone ($R\gg R'$) are:

Circular cylinder,

$$G_1(r,r')\sim\frac{e^{ikR-ikz'\cos\theta}}{4\pi R}\sum_{m=-\infty}^{\infty}\frac{e^{im(\phi-\phi')-im(\pi/2)}}{H_m^{(1)}(ka\sin\theta)}\times[H_m^{(1)}(ka\sin\theta)J_m(k\rho'\sin\theta)-H_m^{(1)}(k\rho'\sin\theta)J_m(ka\sin\theta)], \quad (15)$$

$$G_2(r,r')\sim\frac{e^{ikR-ikz'\cos\theta}}{4\pi R}\sum_{m=-\infty}^{\infty}\frac{e^{im(\phi-\phi')-im(\pi/2)}}{[(\partial/\partial\rho)H_m^{(1)}(k\rho\sin\theta)]_{\rho=a}}\times\begin{Bmatrix}(\partial/\partial\rho)[H_m^{(1)}(k\rho\sin\theta)]_{\rho=a}J_m(k\rho'\sin\theta)\\-(\partial/\partial\rho)[J_m(k\rho\sin\theta)]_{\rho=a}H_m^{(1)}(k\rho'\sin\theta)\end{Bmatrix}. \quad (16)$$

Elliptic cylinder,

$$G_1(r,r')\sim e^{ikR-ikz'\cos\theta}/(2\pi)^{\frac{1}{2}}R\sum_{l=0}^{\infty}[e^{-il(\pi/2)}/Ne,\,o_l]\times[Se,\,o_l(D,v)Se,\,o_l(D,v')/He,\,o_l^{(1)}(D,a)]\times[He,\,o_l^{(1)}(D,a)Je,\,o_l(D,u')-Je,\,o_l(D,a)He,\,o_l^{(1)}(D,u')], \quad (17)$$

$$G_2(r, r') \sim e^{ikR - ikz'\cos\theta}/(2\pi)^{\frac{1}{2}}R$$

$$\times \sum_{l=0}^{\infty} \frac{e^{-il(\pi/2)}}{Ne, o_l} \frac{Se, o_l(D, v) Se, o_l(D, v')}{(\partial/\partial u)[He, o_l^{(1)}(D, u)]_{u=a}}$$

$$\times \{(\partial/\partial u)[He, o_l^{(1)}(D, u)]_{u=a} Je, o_l(D, u')$$

$$- He, o_l^{(1)}(D, u')(\partial/\partial u)[Je, o_l(D, u)]_{u=a}\}, \quad (18)$$

in which $D = \frac{1}{2}kd \sin\theta$.

FAR ZONE FIELDS

General expressions for the far zone field of electric dipole radiators in the presence of a perfectly conducting cylinder are found by substituting the appropriate form of the Green's function in Eqs. (5) and (6) and performing the integration over the current and charge distributions of the dipoles. The latter integration is straightforward, and it is thus possible to treat any distribution of dipoles. To illustrate the application of this formulation the two orientations shown in Figs. 1 and 2 will be considered assuming that the dipoles are of negligible dimensions ($\ll\lambda$).

A. Dipole Parallel to the Axis of the Cylinder (Fig. 1)

In the problem of a dipole carrying a total current, J, and oriented parallel to the axis of the cylinder, Eq. (5) is used to obtain the z component of electric field, which in the far zone is simply related to the θ component by

$$E_z = \sin\theta E_\theta.$$

(In the evaluation of integrals in Eq. (5), the following result which depends upon the equation of continuity of charge and current of the dipole, is utilized

$$\int [J(r')e^{-ikz'\cos\theta} + (\omega/ik^2)\rho(r')(\partial/\partial z')e^{-ikz'\cos\theta}]dv'$$

$$= Jh[1 - \cos^2\theta] = Jh \sin^2\theta.$$

The integration is performed over the volume of a dipole of length, h.) The results are:[7]

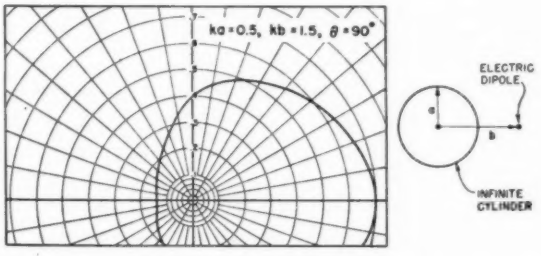


FIG. 6. E_θ *versus* ϕ for an electric dipole parallel to the axis of an infinite circular cylinder.

[7] In a previous report (W. S. Lucke, *Electric Dipoles in the Presence of Elliptic and Circular Cylinders*, Technical Report No. 1, Contract AF 19(122)78, Stanford Research Institute), a result was obtained for this problem which differed from Carter's result (see reference 1). This discrepancy has been corrected in this paper by taking into account the charge density of the dipole which was neglected in the previous report.

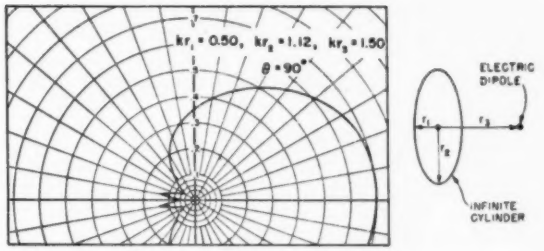


FIG. 7. E_θ *versus* ϕ for an electric dipole parallel to the axis of an infinite elliptic cylinder.

Circular cylinder,

$$E_\theta = (E_z/\sin\theta) \propto (Je^{ikR} \sin\theta/R)$$

$$\times \sum_{m=-\infty}^{\infty} \exp\left\{im\left(\phi - \frac{\pi}{2}\right)\right.$$

$$\left. \times \frac{[H_m^{(1)}(A)J_m(B) - H_m^{(1)}(B)J_m(A)]}{H_m^{(1)}(A)}\right\}, \quad (19)$$

in which $A = ka \sin\theta$ and $B = kb \sin\theta$. The radius of the

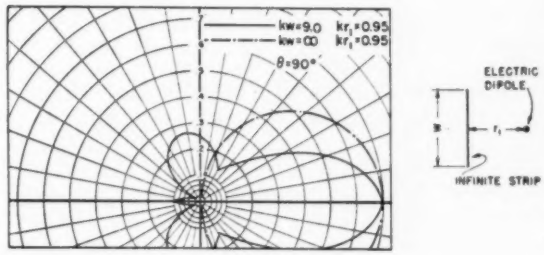


FIG. 8. E_θ *versus* ϕ for an electric dipole parallel to an infinite strip.

cylinder is a, and the coordinates of the dipole are $\rho' = b$, $\phi' = 0$, and $z' = 0$.

Elliptic cylinder,

$$E_\theta = (E_z/\sin\theta) \propto (Je^{ikR} \sin\theta/R)$$

$$\times \sum_{l=0}^{\infty} e^{-il(\pi/2)} \frac{Se, o_l(D, v) Se, o_l(D, v')}{Ne, o_l He, o_l^{(1)}(D, a)}$$

$$\times \left\{ \begin{matrix} Je, o_l(D, b) He, o_l^{(1)}(D, a) \\ -Je, o_l(D, a) He, o_l^{(1)}(D, b) \end{matrix} \right\}, \quad (20)$$

in which $D = (kd/2) \sin\theta$. The cylinder is determined by $u = a$ and the coordinates of the dipole are $u' = b$, $\phi' = 0$, and $z' = 0$.

Examples of the application of Eqs. (19) and (20) are shown in Figs. 6, 7, and 8. Figs. 6 and 7 show the ϕ dependence of E_θ in the plane $\theta = 90°$ for a dipole near a circular and elliptic cylinder respectively. Figure 8 shows the ϕ dependence of E_θ in the plane $\theta = 90°$ for a dipole near an infinitely thin, infinitely long strip which

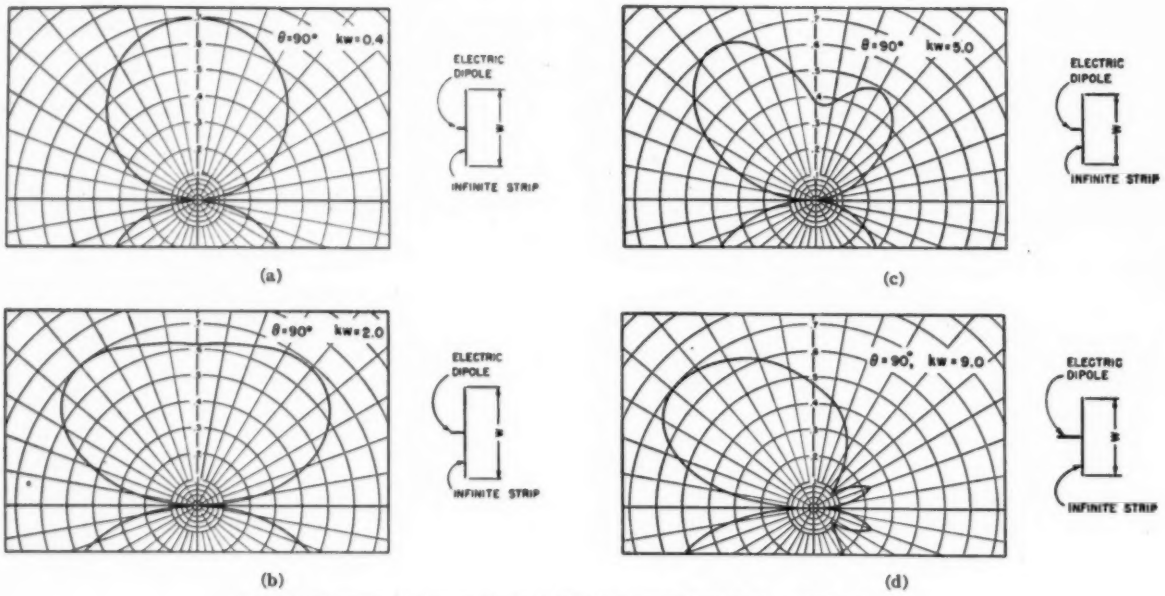


FIG. 9. H_z *versus* ϕ for an electric dipole perpendicular to an infinite strip.

was treated as a degenerate ellipse. For comparison purposes, the field of a dipole correspondingly located with respect to an infinite ground plane has been included.

B. Dipole in a Plane Perpendicular to the Axis of the Cylinder (Fig. 2)

In order to treat the problem of a dipole carrying a total current, J, and located in a plane perpendicular to the axis of a cylinder, use is made of Eq. (6) to obtain the far zone components of magnetic field. In the particular orientations of the dipole along the ρ direction in the circular case and along the u direction in the elliptic case the results are:

Circular cylinder,

$$H_z(r) \propto \frac{Je^{ikR}}{R} \sum_{m=-\infty}^{\infty} \frac{me^{im(\phi-\pi/2)}}{(\partial/\partial A)[H_m^{(1)}(A)]} \times \{J_m(B)(\partial/\partial A)[H_m^{(1)}(A)] - H_m^{(1)}(B)(\partial/\partial A)[J_m(A)]\}, \quad (21)$$

in which $A=ka\sin\theta$ and $B=kb\sin\theta$, the radius of the cylinder is $\rho=a$, and the coordinates of the dipole are $\rho'=b$, $\phi'=0$, and $z'=0$;

Elliptic cylinder,

$$H_z(r) \propto \frac{Je^{ikR}}{R} \sum_{l=0}^{\infty} \frac{e^{-il(\pi/2)} Se,\, o_l(D, v)(\partial/\partial\phi')[Se,\, o_l(D, \cos\phi')]_{\phi'=0}}{Ne,\, o_l(\partial/\partial a)[He,\, o_l^{(1)}(D, a)]} \times \{Je,\, o_l(D, b)(\partial/\partial a)[He,\, o_l^{(1)}(D, a)] - He,\, o_l^{(1)}(D, b)(\partial/\partial a)[Je,\, o_l(D, a)]\}, \quad (22)$$

in which $D=(kd/2)\sin\theta$, the cylinder is determined by $u=a$, and the coordinates of the dipole are $u'=b$, $\phi'=0$, and $z'=0$.

The z component of the electric field may be obtained by similar means and together with the z component of the magnetic field completely determine the far zone field. To illustrate these results, the z component of the far zone magnetic field of a dipole very close to an infinite strip (a degenerate ellipse) is shown in Fig. 9 for strips of various widths. This is a form of the problem of a finite ground plane and illustrates the experimentally verified upward tilt of the maximum of radiation above a finite ground plane. Figure 10 is a plot of far zone field of a dipole in the presence of an elliptic cylinder for a particular choice of parameters.

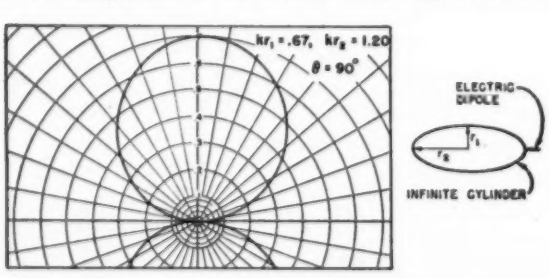


FIG. 10. H_z *versus* ϕ for an electric dipole perpendicular to the surface of an ellipse.

ACKNOWLEDGMENTS

The author wishes to acknowledge the helpful suggestions of Dr. J. V. N. Granger, Supervisor of the

Aircraft Radio Systems Laboratory, Stanford Research Institute, who instigated this research. The aid of Miss Mary Lee Richardson, who performed the computations, is also gratefully acknowledged.

APPENDIX

The derivation of Eqs. (5) and (6) from Eqs. (1) and (2) is based upon the vanishing of the integrations over the sphere at infinity ($R'\to\infty$ and the normal n in the R' direction). This may be demonstrated by noting that from the asymptotic relation between $\mathbf{E}(r')$ and $\mathbf{H}(r')$ for a source $\mathbf{J}(r)$ such that $r'\gg r$ given by

$$n\times\mathbf{E}(r')\sim(\mu/\epsilon)^{\frac{1}{2}}\mathbf{H}(r'), \tag{A-1}$$

and from the asymptotic form of the Green's function given by (See Eqs. 8–11)

$$G(r,r')\sim e^{ikR'}/R', \tag{A-2}$$

it follows that

$$\lim_{R'\to\infty} R'\{i\omega\mu[n\times\mathbf{H}(r')]G(r,r') + [n\times\mathbf{E}(r')]\times n(\partial/\partial R')G(r,r')\}=0, \tag{A-3}$$

and

$$\lim_{R'\to\infty} R'\{i\omega\epsilon[n\times\mathbf{E}(r')]G(r,r') - [n\times\mathbf{H}(r')]\times n(\partial/\partial R')G(r,r')\}=0. \tag{A-4}$$

Substitution of Eqs. (A-3) and (A-4) and the relations

$$n\cdot\mathbf{E}(r')=n\cdot\mathbf{H}(r')=0, \quad R'\to\infty,$$

in Eqs. (1) and (2) yields a vanishing contribution for the integration over the sphere at infinity.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Applications of Potential Theory to the Design of Linear Arrays*

THOMAS T. TAYLOR
Hughes Aircraft Company, Culver City, California

AND

JOHN R. WHINNERY
University of California, Berkeley, California

(Received July 12, 1950)

The polynomial formulation of the linear array problem introduced by Schelkunoff is studied from the point of view of potential theory. An electrical analogy results which is not only very helpful in qualitative thinking about the problem, but which also makes possible the construction of an analog computer for the synthesis of arrays to produce a desired radiation pattern, and for the evaluation of existing arrays. Such a computer has been built, and tested on known patterns with remarkable success.

Several theorems concerning the effect of the roots of the polynomial on the magnitude and phase of the space factor of an array are given. Some of the most important consequences of these are the lack of uniqueness in the array synthesis problem, and the evaluation of known methods of synthesis which consider only constant-phase patterns.

A method of analysis is sketched which in principle does the same thing as the computer, and the difficulties in applying this are described. It is concluded that the computer would be preferable for a wide range of practical problems.

I. INTRODUCTION

AN antenna array may be defined as a series of identical and identically oriented radiating elements whose respective current distributions are geometrically similar. If $A_0, A_1, A_2, \cdots A_{N-1}$ are complex coefficients to which the currents at corresponding points in the several elements are respectively proportional, the power flux pattern in the far zone may be written:

$$\Psi=\Psi_0\left|\sum_{n=0}^{N-1} A_n \exp[ik(\bar{p}_n\cdot\bar{a})]\right|^2. \tag{1}$$

Here Ψ_0 is the power flux pattern of a single element and is known as the element factor. The summation is known as the space factor and contains the excitation coefficients A_n, the respective position vectors $\bar{p}_n$ of the elements, a unit vector $\bar{a}$ which points toward the observer, and the propagation constant k which is equal to $2\pi/\lambda$. In most arrays the space factor, which will be denoted S, is primarily responsible for the directivity of the system; hence, it is the quantity upon which most attention is focused. The remainder of this paper will consider the space factor of a particular type of array, namely, a linear array of equally spaced elements. The discussion will be especially concerned with the problem of designing linear arrays such that the function $|S|$ fulfills initially specified requirements.

II. LINEAR ARRAY THEORY

We wish to make use of the polynomial formulation of the linear array problem discussed in detail by Schelkunoff,[1] so this will be reviewed briefly. Consider

* The work described in this article was performed at Hughes Aircraft Company and was made possible by the sponsorship of the United States Air Forces under Contract W33-038-ac-18205.

[1] S. A. Schelkunoff, Bell System Tech. J. **22**, 80–107 (January 1943).

a linear array whose N elements are situated at positions x_0, x_0+d, $\cdots$ $x_0+(N-1)d$ upon the x axis of a rectangular coordinate system (see Fig. 1). The space factor is given by:

$$S=e^{ikx_0\cos\alpha}[A_0+A_1e^{ikd\cos\alpha}+A_2e^{i2kd\cos\alpha}+\cdots+A_{N-1}e^{i(N-1)kd\cos\alpha}], \tag{2}$$

where α is the angle between the observer's direction and the axis of the array. If the quantity $kd\cos\alpha$ is denoted ψ, a simplified expression is obtained:

$$S=e^{i(x_0/d)\psi}\sum_{n=0}^{N-1}A_ne^{in\psi}. \tag{3}$$

The relationship among ψ, kd, and α is the geometrical one illustrated in Fig. 2. Note that $|S|$ is periodic in ψ, repeating itself at intervals of 2π. The value of d in comparison with a wavelength, that is, kd, determines the size of the semicircle in Fig. 2. This in turn determines how much of the function $|S|$ is translated into

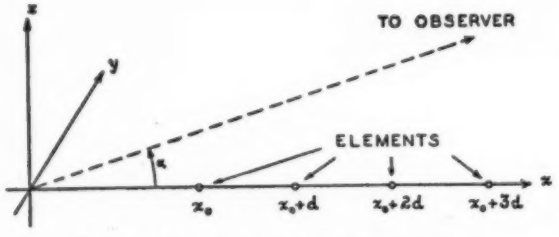


FIG. 1. Linear array and coordinate system.

α-space, that is, how much of this function is visible to an observer who actually travels around the array and examines its pattern. If the elements are spaced a half-wavelength, for example, exactly one period of $|S|$ is visible to such an observer. By choosing the value of kd it is possible to have a pattern in which but a small fraction of one period of $|S|$ is visible, or one in which many periods are visible. The consideration of all such cases is outside the scope of this paper, but from the practical point of view 2π is usually considered an upper bound for kd in crossfire (broadside) arrays and π is such a bound in endfire arrays because it is undesirable to have more than one main lobe present in the visible portion of the pattern.

The technique of using a small value of kd to obtain a visible range small in comparison with a period of $|S|$ and then adjusting $|S|$ so that it exhibits considerably more directivity within this range than without it, is known as "super-gaining" the array. Hansen and Woodyard[2] have shown that the gain of a uniform endfire array can be enhanced by a proper choice of the progressive phase shift of the excitation. This amounts to a translation of the space factor in the coordinate ψ

[2] W. W. Hansen and J. R. Woodyard, Proc. Inst. Radio Engrs. **26**, 333–345 (March, 1938).

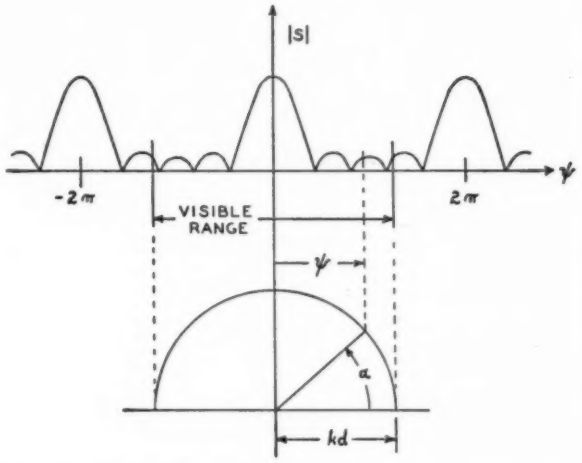


FIG. 2. Demonstrating the relation between ψ-space and α-space for a given kd.

until the maximum of the main lobe is actually outside the visible range, and is a mild form of super-gaining. Schelkunoff[1] has demonstrated that a strategic placement of the zeros of $|S|$ within the visible range can theoretically yield an endfire array of surprisingly high gain. A moderate advantage (about 3 db gain) is obtainable in practice by these techniques. A similar strategy is theoretically applicable to crossfire (broadside) arrays also, but the techniques very quickly give rise to high "invisible" lobes which represent large quantities of stored energy in comparison with the radiated power. Ramo and Taylor[3] and Chu[4] showed from the standpoint of basic electromagnetic theory that no hope for the significant improvement of high gain crossfire arrays by super-gaining can be entertained in practice.

Since super-gaining is not usually recommended, there is rarely any reason to use a small value of kd. In fact, it is desirable to approach the upper bound in selecting a value of kd, since this results in fewer elements. Thus, the visible range will usually cover, or nearly cover, an entire period of $|S|$ and the problem awaiting solution is that of finding the appropriate excitation coefficients A_0, A_1, A_2, $\cdots$ A_{N-1} which will yield the required $|S|$ over a 2π-range in the variable ψ. The feature which presents difficulty is that only the magnitude is specified. If $S(\psi)$ in magnitude and phase was given, the solution would be immediately available by Fourier analysis. With $|S|$ specified, the problem is more subtle and the solution is not, in general, unique. Eaton, Eyges, and Macfarlane[5] have recognized this problem and have given some suggestions as to its solution. Their method is different from, but related to, the methods to be discussed here.

[3] T. T. Taylor, Proc. Inst. Radio Engrs. **36**, 1135 (September, 1948).

[4] L. J. Chu, J. Appl. Phys. **19**, 1163 (1948).

[5] S. Silver and others, *Microwave Antenna Theory and Design* (McGraw-Hill Book Company, Inc., New York, 1949), pp. 258 to 284.

III. SOLUTION OF THE LINEAR ARRAY PROBLEM BY AN ELECTRICAL ANALOG

To continue with the development of the polynomial approach in which S is regarded as a function of the complex variable, consider the expression:

$$S=z^{(x_0/d)}\sum_{n=0}^{N-1}A_nz^n. \tag{4}$$

Here $z=x+iy=re^{i\psi}$. Observe that (4) reduces to (3) upon the unit circle $r=1$. If we take the logarithm of (4), magnitude and phase are separated, becoming related respectively to the real and imaginary parts of the new function, $\log S$:

$$\log S=\log|S|+i\arg S. \tag{5}$$

It is here that the analog comes into play, for $\log S$ is a function of the form $f(z)=u(z)+iv(z)$ and it is a well-known fact that u and v are Laplacian in regions where $f(z)$ is analytic. The electric potential, U, in a two-dimensional isotropic conductor is also Laplacian and forms a most convenient analog. It is only necessary to find a system of sources and sinks for the singularities of $u(z)$, that is, $\log|S|$, in order to produce a potential everywhere proportional to this quantity. The problem is similar to that described by Hansen and Lundstrom[6] for network functions.

Let us now reconsider Eq. (4) with the polynomial replaced by the product of its $N-1$ linear factors:

$$S=A_{N-1}z^{(x_0/d)}(z-z_1)(z-z_2)\cdots(z-z_{N-1}). \tag{6}$$

Evidently the real part of $f(z)$, $\log|S|$, is as follows:

$$u=\log|A_{N-1}|+(x_0/d)\log|z|+\sum_{n=1}^{N-1}\log|z-z_n|. \tag{7}$$

Consider one term of the above as $u_1=\log|z-z_1|$. If a current I is introduced into the two-dimensional conductor at the position denoted by the complex number z_1 and removed at infinity, it will be found that the electric potential at point z in this type of field is given by:

$$U_1=\frac{-IR_s}{2\pi}[\log|z-z_1|+K_1], \tag{8}$$

where R_s is the per-square resistance of the conducting medium and K_1 is an arbitrary constant. The case is similar for other sources of constant current I at $z_2\cdots z_{N-1}$, so that a superposition of such potential fields will evidently yield an analogy to the corresponding terms in Eq. (7). To simulate properly the term $(x_0/d)\log|z|$ a current of value x_0I/d is introduced at the origin. Addingthis field also, we obtain as a total potential:

$$U=\frac{-IR_s}{2\pi}\left[(x_0/d)\log|z|+\sum_{n=1}^{N-1}\log|z-z_n|+K\right], \tag{9}$$

where K denotes the sum of all arbitrary constants. This K corresponds to $\log A_{N-1}$, and so may be set equal to zero, for in the final array only the relative values of the excitation coefficients are significant and any one of these, e.g., A_{N-1}, may be set equal to unity, whereupon its logarithm becomes zero. With the aid of these concepts, we may write:

$$U=(-IR_s/2\pi)\log|S|, \tag{10}$$

and the analogy is complete and exact. In particular, if U is measured upon the unit circle it will be found to be proportional to $\log|S|$ as a function of ψ for the corresponding linear array. That is, the potential measured on the model is proportional to the power pattern of the antenna array measured on a logarithmic (decibel) scale. The physical equipment required is sketched in Fig. 3. The use of this equipment in solving the linear array design problem consists in manual arrangement of the root positions (points of entry of the currents) until the relative variation of the quantity U as measured upon the unit circle, preferably by an automatic device, is as desired. The various currents which act as the sources of the field are maintained constant during this process, and all are equal to one another with the exception of the current entering at the origin, which is different from the rest by a factor of x_0/d. The quantity x_0 is simply the displacement of the first element of the array from the origin of physical coordinates (see Fig. 1) and should have no effect upon the magnitude of the space factor. This is true, for any value of current introduced at the origin of the complex plane (see Fig. 3) produces no potential variation upon the unit circle.

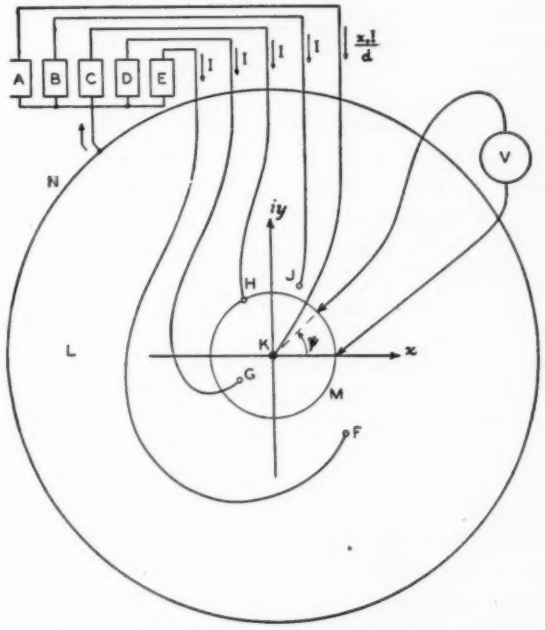


FIG. 3. Illustration of the electrical analog for the linear array problem. *A*, *B*, *C*, *D*, *E*—constant current generators. *F*, *G*, *H*, *J*—roots. *K*—root at the origin. *N*—large circular electrode corresponding to infinity. *M*—the unit circle. *V*—voltmeter.

[6] W. W. Hansen and O. C. Lundstrom, Proc. Inst. Radio Engrs. 33, 528–534 (August, 1945).

Thus x_0/d is arbitrary as far as $\log|S|$ is concerned. For example, all the current entering the field at the root positions can be removed at the origin, a device which is equivalent to setting x_0/d equal to $-(N-1)$ and which obviates the necessity of collecting current at infinity. The effect of x_0/d upon the phase of S will be discussed in later paragraphs.

It is not difficult to become skilled in arranging the root positions to obtain a desired relative variation of U around the unit circle. Considerable background for this already exists in our knowledge of the behavior of potential fields in general. Thus the most significant effect of any root is localized to its immediate neighborhood and we may make adjustments upon one portion of the pattern without seriously disturbing other portions. Roots near the unit circle have the greatest effect upon the pattern, and roots actually on the unit circle correspond to a negative infinity in $\log|S|$ or a null in the radiation. Complicated patterns obviously correspond to high directivity and require more roots than do simple patterns; the synthesis of a particular pattern should first be attempted with a few roots whereupon others may be added if necessary. When the process of synthesis is complete, the root positions $z_1, z_2, \cdots z_{N-1}$ are recorded as complex numbers and the product $(z-z_1)(z-z_2)\cdots(z-z_{N-1})$ is expanded to give the coefficients $A_0\cdots A_{N-1}$ which are the desired relative currents of the array.

To help evaluate the usefulness of these techniques, a small-scale experimental model of the analog computer was constructed. This model consists of a network of resistors arranged in square meshes forming a two-dimensional resistive medium which corresponds by transformation to the interior of the unit circle in the complex plane. (It was found that an imaging process could be applied which permitted the use of only the interior of the unit circle.) Convenient values of resistance, current, and voltage could be used throughout the model. Data obtained were reliable and repeatable. The problem of accuracy was considered in detail, especially with regard to the question of the use of a network rather than a continuous medium. Errors arising from this cause are totally negligible in areas of the plane remote from roots but do appear in the neighborhood of such points. At the root itself, an infinite error exists because the point in question cannot assume a potential of minus infinity, although its potential does reach a very low value. At mesh points adjecent to the root, the greatest finite error exists; this error was found to be about one-third of a decibel for a square mesh and was of the opposite sign to that expected. Additional research showed that this result could be theoretically explained, and that the error could be reduced to about one twenty-fifth of a decibel by the use of a network with triangular meshes rather than one with square meshes.

It is felt that the model computer gave adequate assurance that the principles here developed are capable of practical utilization and that the accuracy obtainable is consistent with the most exacting engineering requirements. A more detailed description of the model will be presented in a later paper.

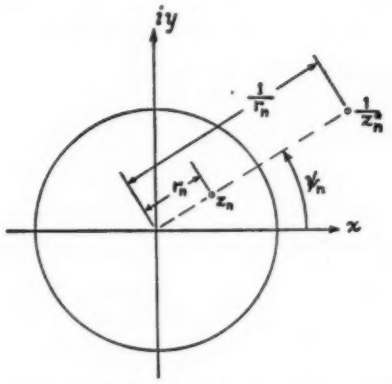


FIG. 4. The image points in the unit circle, z_n and $1/z_n^*$.

IV. THEOREMS RELATING TO THE MAGNITUDE OF THE SPACE FACTOR

Since the basic concepts associated with the solution of the linear array problem by an electrical analog have been presented, we may proceed to the consideration of some important auxiliary theorems. Many of these are known to workers in arrays,[5] and others are more or less evident from corresponding theorems of network synthesis, but they are presented here because they are most important to the intelligent use of the analog described. The first of these concerns the question of image points in the unit circle. For any complex number z_n, there is defined an image point $1/z_n^*$ which occurs at the same polar angle as z_n but has a magnitude which is the reciprocal of the magnitude of z_n (see Fig. 4). We now state the following theorem:

Theorem I: If any or several of the zeros of the polynomial portion of $S(z)$, hereinafter simply called the roots of $S(z)$, are moved to their respective image points, the relative variation of $\log|S|$ on the unit circle is unchanged

As proof of this theorem, consider two space factors S and S_1 defined as follows:

$$S=S_0(z-z_n) \tag{11}$$

$$S_1=S_0(z-1/z_n^*). \tag{12}$$

Here S_0 is an abbreviation for all the factors of S except $(z-z_n)$, and S_1 differs from S only in that z_n has been replaced by its image point. We will express z_n as $r_n e^{i\psi_n}$ and proceed as follows:

$$\log|S|=\tfrac{1}{2}\log SS^*=\tfrac{1}{2}\log S_0S_0^* + \tfrac{1}{2}\log[(e^{i\psi}-r_ne^{i\psi_n})(e^{-i\psi}-r_ne^{-i\psi_n})] \tag{13}$$

$$\log|S|=\tfrac{1}{2}\log S_0S_0^* + \tfrac{1}{2}\log[1-2r_n\cos(\psi-\psi_n)+r_n^2]. \tag{14}$$

Similarly for S_1:

$$\log|S_1| = \tfrac{1}{2}\log S_0 S_0^* + \tfrac{1}{2}\log\left[1 - \frac{2}{r_n}\cos(\psi-\psi_n) + \frac{1}{r_n^2}\right]. \quad (15)$$

Evidently:

$$\log|S_1| = \log|S| - \log r_n. \quad (16)$$

Thus the two quantities differ by a constant, which means that their relative variation around the unit circle is the same. The theorem has been proved for a particular root, z_n, and can be extended to cover the case in which several roots are moved to their image points.

We are now faced with a question of uniqueness of solution, for if a root of $S(z)$ is moved to its image point, the relative pattern of the array is unchanged, yet the series of excitation coefficients is substantially modified. Evidently there are cases in which more than one array can produce the same relative power pattern. We will attempt to organize these cases, beginning with a statement of the following theorem.

Theorem II: If all the roots of $S(z)$ lie on the unit circle, the solution to the array problem is unique

This is evident, for roots on the unit circle already coincide with their image points, and no movement of roots in the sense of Theorem I is possible. For such arrays, $|S|$ goes to zero $N-1$ times in a 2π range of ψ. High gain arrays are usually of this type, for a high gain array is one which gives strong radiation in the neighborhood of one value of ψ and as little radiation as possible for all other values. There is no better way to insure that the average value of $|S|$ be small over a particular range than to make $|S|$ go to zero as many times as possible within that range. Uniform arrays and Dolph[7] arrays are particular examples. Arrays which do not come within the scope of Theorem II are also important, and are more interesting from the point of view of this paper. These are usually required when a specially shaped pattern is specified, and since such arrays have roots not on the unit circle, special uniqueness theorems must be developed for them. The more elementary of these may be stated as follows.

Theorem IIIa: If, in a linear array, q_1 of the roots of $S(z)$ lies off the unit circle, and if none of these are multiple roots or roots lying at the image points of one another, there are 2^{q_1} different series of excitation coefficients which will produce the same relative variation of $\log|S|$ upon the unit circle

This theorem follows at once from Theorem I, for in moving any one of the q_1 roots to its image point a new array is formed, but the relative power pattern remains the same, and the total number of permutations is 2^{q_1}.

[7] C. L. Dolph, Proc. Inst. Radio Engrs. **34**, 335–348 (June, 1946).

As a simple example take a three-element array with the two roots at $\frac{1}{2}$ and $-\frac{1}{2}$. The $2^2=4$ possible arrays that produce the same relative power pattern are:

$$\left.\begin{aligned} S_1(z) &= (z-\tfrac{1}{2})(z+\tfrac{1}{2}) = -\tfrac{1}{4}+0+z^2 \\ S_2(z) &= (z-\tfrac{1}{2})(z+2) = -1+\tfrac{3}{2}z+z^2 \\ S_3(z) &= (z-2)(z+\tfrac{1}{2}) = -1-\tfrac{3}{2}z+z^2 \\ S_4(z) &= (z-2)(z+2) = -4+0+z^2 \end{aligned}\right\}. \quad (17)$$

The coefficients of the polynomials are the different possible current excitations in the three elements. The relative power patterns are all the same, since:

$$4S_1S_1^* = S_2S_2^* = S_3S_3^* = \tfrac{1}{4}S_4S_4^* = \frac{17}{4} - 2\cos 2\psi. \quad (18)$$

This lack of uniqueness is especially impressive for a large array. For example, a twenty-element array with no zeros on the unit circle could have $2^{19}=524,288$ different current distributions producing the same relative power pattern. We now advance to a more general Theorem.

Theorem IIIb: If, in a linear array, the roots of $S(z)$ not on the unit circle comprise q_1 single roots, q_2 double roots, $\cdots$ q_s roots of multiplicity s, there are then $2^{q_1}3^{q_2}4^{q_3}\cdots(s+1)^{q_s}$ different series of excitation coefficients which will produce the same relative variation of $\log|S|$ upon the unit circle. For the purpose of this theorem, two roots lying at the image points of one another are equivalent to a double root

The revision of Theorem IIIa to give Theorem IIIb depends upon the fact that a double root gives three essentially different arrangements rather than the 2^2 or four arrangements that would follow if there were two separate single roots, for two, one, or none of the pair may be at the original point with the remainder at the image point. The situation would not be different if one started with two single roots at the image points of one another. In general, for a root of multiplicity s there are $s+1$ arrangements. If there are q_s roots of multiplicity s, it is possible to choose any of the $s+1$ arrangements associated with the first such root, then (independently) any of the $s+1$ arrangements associated with the second such root, etc. The total number of arrangements becomes $(s+1)^{q_s}$; hence, Theorem IIIb is as stated. Note that if there are no multiple roots (in the extended sense) off the unit circle, Theorem IIIb reduces to Theorem IIIa.

An additional theorem pertaining to the "taper" of the coefficients may be useful, especially when the engineer must decide among alternative sequences of coefficients some of which may be more suitable than others to the plan envisioned for the feeding of the elements. This theorem is as follows.

Theorem IV: Of all the possible root distributions that produce the same relative power pattern, the one with all roots which are not on the unit

circle lying inside the unit circle has the greatest ratio of last-coefficient to first-coefficient magnitudes; the distribution with all such roots lying outside the unit circle has the least such ratio. If the root distribution is such that the magnitude of the product of the roots is unity (as, for example, when all the roots lie on the unit circle), the last and first coefficients are equal in magnitude

This is evident from a study of the formation of the product:

$$S(z)=(z-z_1)(z-z_2)(z-z_3)\cdots(z-z_{N-1}). \quad (19)$$

Here the coefficient of z^{N-1}, i.e., the last coefficient, is always unity. The first coefficient is simply the product of the negatives of the roots,

$$A_0=\prod_{n=1}^{N-1}(-z_n) \quad (20)$$

and its magnitude is least when all the z_n not on the unit circle are within it and greatest when these z_n are outside. The magnitudes of the coefficients do not always increase or decrease monotonically in the respective cases. Nevertheless, a limited experience with such calculations has shown that a general tendency in this direction exists. Attention is called to it becanse of its importance to the practical problem of feeding the array.

V. THEOREMS RELATING TO THE PHASE OF THE SPACE FACTOR

Let us now consider the phase surface $v(x, y)$. We must go back to Eq. (6) and write:

$$v=\arg S=(x_0/d)\psi+\sum_{n=1}^{N-1}\arg(z-z_n). \quad (21)$$

Here again A_{N-1} has been set equal to unity so that its argument becomes zero. Considering terms separately, let

$$v_1=\arg(z-z_1) \quad (22)$$

and furthermore, let us station an observer at the variable point z for the purpose of exploring the variation of v_1 in all parts of the plane. This observer will now find himself upon a sort of helical staircase the surface of which is discontinuous at z_1 (the central supporting member of the staircase) but otherwise increases smoothly by 2π for every complete revolution in the positive direction. This is to be expected, for the singularities of $f(z)=\log S$ are branch points, and $f(z)$ is multivalued, the multiplicity of values being found only in the imaginary part. If we were to change the analogy, simulating v_1 by an electric potential V_1, we would be forced to cut the conducting plane from infinity to the point z_1, and between the shores of this "branch cut" we would establish a distributed electromotive force, B. Then V_1 would be given by:

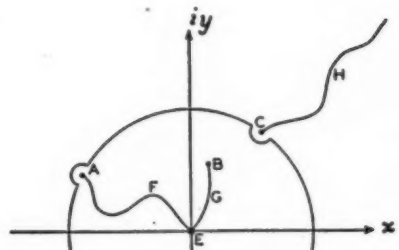


FIG. 5. Illustration of branch cuts in the z plane. A, B, C, D, E—roots of $S(z)$ and hence branch points of $\log S(z)$. F, G, H, J—branch cuts. K—indented contour which approaches the unit circle.

$$V_1=\frac{B}{2\pi}\arg(z-z_1)=\frac{Bv_1}{2\pi}. \quad (23)$$

The cut plane which we now possess simulates only one branch of v_1, which is all we could hope to accomplish in a physical system of this kind. In the case of the singularity at the origin, we have a helical staircase whose pitch is controlled by the quantity x_0/d, and may accordingly have any value, positive, zero, or negative. It will be helpful to note that if x_0/d is a negative integer, m, the branch cuts from m of the roots need not go to infinity, but can terminate at the origin. If $x_0/d=-(N-1)$, no branch cuts go to infinity—a case exactly analogous to that already discussed in connection with the function u.

Again, the profile of v upon the circumference of the unit circle corresponds exactly to the phase function of S in radians. We should note that a passage of a branch cut through the circle does not constitute a discontinuity in $\arg S$. Branch cuts are mental crutches; the true picture of v is the totality of all its branches—that is, an infinite number of smoothly connected "levels" which an observer visits successively if he continues to encircle one of the singularities. A branch point (root) exactly on the unit circle, however, does introduce a discontinuity into $\arg S$, for in this case the profile of v will include an abrupt step of magnitude π. We therefore state the following theorem.

Theorem V: At every single root of $S(z)$ on the unit circle, $\arg S$ changes by π

This is an important theorem stating what many readers have long suspected or known—namely, that the phase of the space factor reverses itself at points where its magnitude goes to zero in the abrupt fashion characteristic of a single root. A more mathematical approach

to this phenomenon may be based upon the observation that the term $\arg(z-z_n)$ changes by π as z, constrained to follow the unit circle, passes through z_n and emerges on the other side. Passage through branch points of the type here encountered may be regarded as the limiting case of passage around such points via the indented contour of Fig. 5 as the radius of the indentations approach zero. We have already discussed the case in which x_0/d is equal to a negative integer, m. We can now see that if a proper choice of m is made, we can arrange the branch cuts so that none of these cross the unit circle. For this case, then, $\arg S$ remains on the same branch and returns to its initial value after ψ has increased through a 2π-range. Specifically, if there are l_1 roots on the unit circle and l_2 roots within the unit circle, x_0/d must lie in the range:

$$-(l_1+l_2)\leq(x_0/d)\leq -l_2 \tag{24}$$

and must be an integer, if $\arg S$ is to remain on the same branch. If none of the roots are on the unit circle, we no longer have the privilege of arbitrarily directing the associated branch cuts inward or outward, and x_0/d must be exactly equal to $-l_2$ for $\arg S$ to remain upon the same branch. In either case, it requires that the origin of physical coordinates be within the array and coincident with one of the elements.

A very important case is the one in which $\arg S$ is a linear function of ψ upon those portions of the unit circle not occupied by roots. We state the following theorem.

Theorem VI: If each of the roots of $S(z)$ lies either on the unit circle or at the image point of another root, $\arg S$ is a linear function of ψ upon all portions of the unit circle except at those points actually occupied by roots. Arrays having this property will be known as arrays with linear phase variation

In order to prove this theorem we shall use expression (6) with $A_{N-1}=1$, and both multiply and divide by the factor:

$$[z^{N-1}(-z_1)(-z_2)(-z_3)\cdots(-z_{N-1})]^{\frac{1}{2}}.$$

Thus we have:

$$S=z^{[(x_0/d)+(N-1)/2]}A_0^{\frac{1}{2}} \times\left[\frac{(z-z_1)}{(-zz_1)^{\frac{1}{2}}}\cdot\frac{(z-z_2)}{(-zz_2)^{\frac{1}{2}}}\cdots\frac{(z-z_{N-1})}{(-zz_{N-1})^{\frac{1}{2}}}\right]. \tag{25}$$

The phase function is then:

$$\arg S=[(x_0/d)+(N-1)/2]\psi +\tfrac{1}{2}\arg A_0+\sum_{n=1}^{N-1}\arg\frac{(z-z_n)}{(-zz_n)^{\frac{1}{2}}}. \tag{26}$$

We will first consider the case where a root, z_n, lies on the unit circle. Then:

$$\frac{(z-z_n)}{(-zz_n)^{\frac{1}{2}}}=\frac{(e^{i\psi}-e^{i\psi_n})e^{-i(\psi+\psi_n)/2}}{i}=2\sin\frac{(\psi-\psi_n)}{2} \tag{27}$$

$$\arg\frac{(z-z_n)}{(-zz_n)^{\frac{1}{2}}}=q\pi. \tag{28}$$

Here q is an integer which either increases or decreases by one each time ψ passes through ψ_n. Thus a factor such as (27) adds a "stair-step" function to $\arg S$, creating the discontinuities. Let us now consider the other possibility, a pair of roots at each other's image points. It can be shown that

$$\frac{(z-z_n)(z-1/z_n^*)}{-z(z_n/z_n^*)^{\frac{1}{2}}}=(r_n+1/r_n)-2\cos(\psi-\psi_n), \tag{29}$$

which is always a non-negative real. Hence, any factor such as (29) involving a pair of roots at image points has no effect upon $\arg S$. The continuous portion of $\arg S$ is then given by:

$$f(\psi)=[(x_0/d)+(N-1)/2]\psi+\tfrac{1}{2}\arg A_0, \tag{30}$$

which is the linear function required by Theorem VI. Evidently, by choosing the origin of physical coordinates such that $x_0/d=-(N-1)/2$, $f(\psi)$ becomes a constant. We may write a theorem to cover this case as follows.

Theorem VII: An array with linear phase variation becomes an array whose space factor is of constant phase if the origin of coordinates is chosen at the physical center of the array

By constant we mean constant between discontinuities, of course. Choosing x_0/d equal to $-(N-1)/2$ is

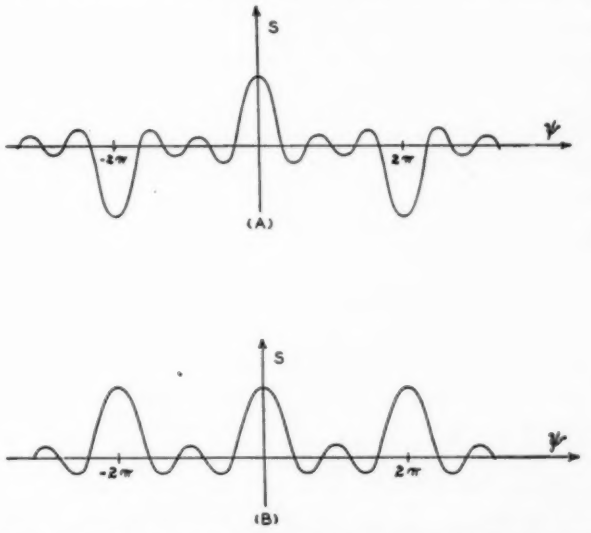


FIG. 6. Comparison of the space factors of a six-element uniform array (*A*), and a five-element uniform array (*B*). Both arrays are centered at the origin.

equivalent to placing the origin of coordinates at the physical center of the array, so the theorem is true. Note that the case where arg$S$ remains upon the same branch as ψ increases is not necessarily the same as the case wherein arg$S$ is a constant between roots. For example, consider an array with a space factor of constant phase between roots and with an even number of elements. The function arg$S$ for such arrays changes by π after each traverse of the unit circle and successive periods in ψ are characterized by a faithful repetition of $|S|$ but by an inversion of phase. See, for example, the space factor of a six-element uniform array contrasted with that of a five-element uniform array, both arrays being centered at the origin (Fig. 6).

VI. METHODS OF SYNTHESIS BASED UPON CONSTANT-PHASE PATTERNS

There are several well-known methods of array synthesis which have been used mainly with a real space factor $S_r(\alpha)$ selected such that $|S_r(\alpha)|$ is equal to the desired $|S(\alpha)|$. The patterns are thus limited to those which are of constant phase between singularities, as in Theorem VII, and the question arises as to the loss of flexibility in limiting the problem in this way.

One method of synthesis which has been described by Berndt,[8] Wolff,[9] and Schelkunoff[1] is based upon the approximation of $S(\alpha)$ by a trigonometric series with a finite number of terms, and this may be produced by a finite array. Thus, given the desired magnitude as a function of α, the phase angle is set equal to zero so that the resulting real $S_r(\alpha)$ can be interpreted as a function of $\psi = kd\cos\alpha$ and expanded as a Fourier series in ψ. A finite number of terms, m, is retained to approximate the pattern to a desired degree of accuracy.

$$S_r(\psi) = c_0 + 2\sum_{q=1}^{m} c_q \cos q\psi + d_q \sin q\psi. \tag{31}$$

This approximating pattern may be obtained with an array of $(2m+1)$ elements if the excitation coefficients of the array are related to the Fourier coefficients as follows:

$$\left.\begin{aligned} A_{m-q} &= c_q + id_q = A^*_{m+q} \\ A_m &= c_0 \end{aligned}\right\}. \tag{32}$$

Then

$$S_r(\psi) = \sum_{q=m}^{1} A_{m-q}e^{-iq\psi} + A_m + \sum_{q=1}^{m} A_{m+q}e^{iq\psi}, \tag{33}$$

which, with the substitution of (32), reduces to (31).

A second method of pattern synthesis[10] utilizes a series of terms of the type $(\sin\psi)/\psi$ to approximate the desired space factor. The excitation function is found by taking the Fourier transform of the space factor to yield a sum of rectangular current blocks of various magnitudes and phases which may be added to give a continuous current distribution, which may in turn be approximated by the currents in a discrete number of elements. This method has been described in the literature for use with constant phase patterns only. However, it is understood that it has also been used with a variation in phase between the terms of the series, with a consequent improvement in the economy of design. This fact suggests that a limitation to constant phase patterns may be uneconomical. The reasons for this may be clarified by means of the preceding section.

Referring to Theorems VI and VII, it is known that constant phase patterns result if all the roots lie on the unit circle, or if all roots off the unit circle are paired at the image points with respect to the unit circle. The origin must also be chosen at the center of the array, but this is no limitation. For high gain arrays, the roots are used to best advantage if placed on the unit circle, so for such arrays there should be no loss in limiting synthesis methods to those which consider constant phase patterns only.

For a shaped pattern which required roots off the unit circle, suppose a synthesis method applied which yielded the best distribution of roots, with all roots inside the unit circle. The pattern could be made exactly constant phase only if all such roots turned out to be double roots so that one from each pair could be moved to the outside image point. This would be unlikely, but if there were a large number of roots, it is probable that for each root another could be found near it (compared with the distance to the unit circle boundary). Such closely neighboring roots could be combined into sets of double roots without doing much violence to the pattern, whereupon a constant phase condition could be obtained by the expedient just discussed. If there are few roots, however, it is not likely that this process could be carried out without destroying much of the smoothness of the pattern. This is particularly true if some of the roots are placed off the unit circle but near it (as illustrated in Fig. 7) in order to shape the low intensity regions.

Summarizing, we believe that there is little lost in limiting oneself to methods of synthesis which consider

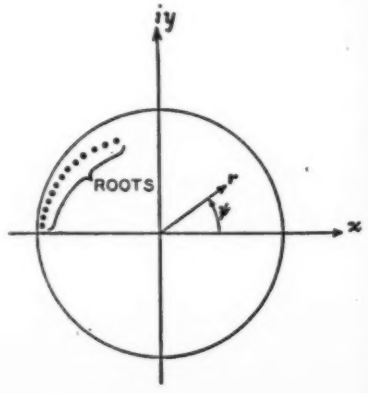


FIG. 7. Demonstrating the shaping of the low intensity part of the pattern by roots near the unit circle.

[8] W. Berndt, Hochfreq. und Elek. 44, 23–28 (July, 1934).
[9] I. Wolff, Proc. Inst. Radio Engrs. 25, 630–643 (May, 1937).
[10] P. M. Woodward, J. Inst. Elec. Engr. Part IIIa, 95, 1554–1558 (1946).

only constant phase patterns when one is concerned with high gain arrays, or with most arrays having a large number of elements. These methods will frequently lead to uneconomical designs when the number of elements in the array to be synthesized is not very great, or if it is desired to shape a particular part of the pattern by roots close to the unit circle. This conclusion is in harmony with Eaton, Eyges, and Macfarlane's comments regarding the generality of methods involving constant phase patterns.[5]

VII. SOLUTION OF THE LINEAR ARRAY PROBLEM BY ANALYSIS

Having carefully considered the potential analog as a means for solving the array design problem, it is desirable to examine analytical methods which can in principle yield the same result.

We will suppose that $u=\log|S|$ has been specified for $0\leq\psi<2\pi$ and that the object is to find at least one and preferably all of the linear arrays which will produce this given function. We will begin by finding a particular one of the desired arrays, namely, the one in which all the roots of $S(z)$ are either on the unit circle or outside it. We will furthermore specify that $x_0/d=0$, thereby making the origin free of singularities. All branch cuts now retreat to infinity, and an observer who follows the indented contour D of Fig. 8 will always remain upon the same branch of $\arg S$. The interior of the unit circle is now a circular region in which $f(z)=\log S$ is analytic, and the function may be written as a Taylor's series about the origin:

$$\log S=\sum_{n=0}^{\infty} C_n z^n. \tag{34}$$

Here C_n is complex, and may be written as follows:

$$C_n=\alpha_n+i\beta_n, \tag{35}$$

where α_n and β_n are real. The real and imaginary parts of $\log S$ are then:

$$u=\sum_{n=0}^{\infty} r^n(\alpha_n\cos n\psi-\beta_n\sin n\psi) \tag{36}$$

$$v=\sum_{n=0}^{\infty} r^n(\alpha_n\sin n\psi+\beta_n\cos n\psi). \tag{37}$$

Of the above functions, u only is specified and this for all ψ from zero to 2π at $r=1$.

If we can find α_n and β_n from this knowledge only, the problem is solved. By Fourier's Theorem we can find all α_n and all β_n except β_0 in the following way (where the Cauchy value of the integral is to be taken if u becomes infinite upon the unit circle):

$$\alpha_0=\frac{1}{2\pi}\int_0^{2\pi} u\,d\psi \tag{38}$$

$$\alpha_n=\frac{1}{\pi}\int_0^{2\pi} u\cos n\psi\,d\psi \qquad (n\neq 0) \tag{39}$$

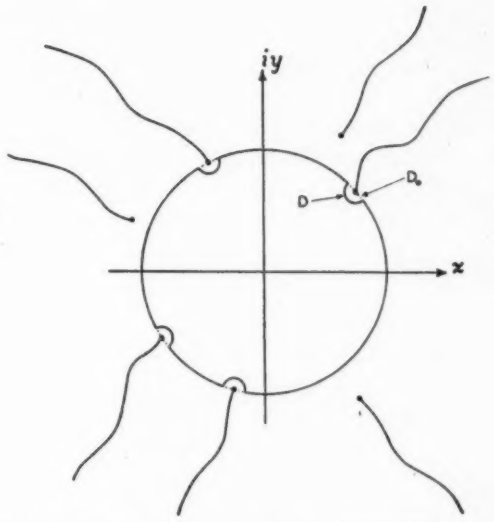


FIG. 8. Root configuration for which $f(z)=\log S(z)$ is analytic in the interior of the unit circle.

$$\beta_n=\frac{-1}{\pi}\int_0^{2\pi} u\sin n\psi\,d\psi. \quad (n\neq 0) \tag{40}$$

The two latter formulas may be combined into the more concise form:

$$C_n=\frac{1}{\pi}\int_0^{2\pi} ue^{-in\psi}d\psi \quad (n\neq 0), \tag{41}$$

which may also be written:

$$C_n=\lim_{D\to D_0}\left[\frac{1}{\pi i}\int_D \frac{u\,dz}{z^{n+1}}\right] \quad (n\neq 0). \tag{42}$$

Here D is the indented contour of Fig. 8, and D_0 is the circumference of the unit circle itself. Notice that C_0 is an exceptional case, and its imaginary part cannot be found. This is no handicap, for C_0 is merely the constant part of $\log S$ (corresponding to a numerical factor in S) and is therefore arbitrary.

If the above process is carried out, we obtain the power series coefficients of $\log S$; the problem remaining is that of finding a similar set of coefficients for S itself, since the latter will actually be the excitation coefficients of the array. These coefficients can be found merely by applying the expansion theorem to the function $e^{\log S}$ as follows:

$$A_m=\frac{1}{m!}\left[\frac{d^m}{dz^m}\exp\left(\left\{\sum_{n=0}^{\infty} C_n z^n\right\}\right)\right]_{z=0}. \tag{43}$$

Equation (43) yields the following for the first few orders:

$$\left.\begin{aligned} A_0&=e^{C_0}\\ A_1&=A_0C_1\\ A_2&=\tfrac{1}{2}[2A_0C_2+A_1C_1]\\ A_3&=\tfrac{1}{3}[3A_0C_3+2A_1C_2+A_2C_1]\\ A_4&=\tfrac{1}{4}[4A_0C_4+3A_1C_3+2A_2C_2+A_3C_1]\\ &\text{etc.}\end{aligned}\right\}. \tag{44}$$

As an example of the method, let us consider the pattern of a two-element uniform array given in magnitude by:

$$u=\log|S|=\log\left|2\cos\frac{\psi}{2}\right|. \tag{45}$$

The Fourier coefficients, obtainable by Eqs. (38) through (42) are:[11]

$$\left.\begin{aligned}&\alpha_0=0\\&\alpha_1=1\\&\alpha_2=-\tfrac{1}{2}\\&\alpha_3=\tfrac{1}{3}\\&\cdots\\&\alpha_n=(-1)^{n-1}/n\quad(n\neq 0)\\&\beta_n=0\quad(n\neq 0)\end{aligned}\right\}. \tag{46}$$

Since the β_n are zero, C_n is real and equal to α_n. We have arbitrarily set β_0 equal to zero. The excitation coefficients of the array are given by (44):

$$\left.\begin{aligned}&A_0=1\\&A_1=A_0\cdot 1=1\\&A_2=\tfrac{1}{2}[2A_0(-\tfrac{1}{2})+A_1]=0\\&A_3=\tfrac{1}{3}[3A_0(\tfrac{1}{3})+2A_1(-\tfrac{1}{2})+A_2]=0\end{aligned}\right\}. \tag{47}$$

A continuation of the process results only in zeros, and we see that the pattern of (45) is indeed obtainable from an array of but two elements in which $A_0=1$ and $A_1=1$. An example involving a root exactly on the unit circle was here chosen in order to emphasize the validity of the method even for such cases.

One of the outstanding features of the practical case, which differentiates it from the foregoing simple illustration, is that the function u, the starting point of the entire process, is generally not specified exactly at all points. Rather there are certain requirements, tolerances, maximum side lobe specifications, etc., applied to u in such a way that many functions could be used with equally good results. The particular functional form chosen for convenience of numerical integration would probably not dip down to minus infinity anywhere, although such singularities might actually be advantageous in terms of economy of design. The sequence of excitation coefficients obtained in a practical problem seldom terminate, and an additional analysis would have to be made to discover how many of these coefficients should be retained. Also, only one of the many possible sets of excitation coefficients would result from the analytical approach just outlined. To obtain the others, it would be necessary to know the roots of $S(z)$ in order that the various permutations created by moving roots to their image points could be investigated. Finding these roots from the excitation coefficients is the difficult problem of solving an algebraic equation of high order. Finally, engineers almost inevitably want to know what would happen to the final solution if the initial requirements were slightly modified and to supply such information by the analytical method would require a reperformance of the entire calculation several times.

The procedures just discussed are to be contrasted with those involving the electrical analog. In using the analog, the operator becomes very much aware of the possibilities of the number of roots he is using, and by gradually adding roots where necessary, he is quite well satisfied that the final pattern could not have been obtained with fewer roots, that is, with an array of fewer elements. He will not be hesitant about putting roots on the unit circle, because the reasons which deterred the analyst will not be present. The answer, when obtained, will consist of simply a list of the roots, and any or all of the associated arrays can be found by the elementary operations of addition and multiplication. So although it is probable that there are situations to which the analytical approach is admirably suited, it is felt that the electrical analog method will be preferable for the general class of engineering problems.

VIII. EVALUATION OF AN EXISTING ARRAY

A problem closely allied with the linear array design problem is that of evaluating an existing array. We here suppose that an array is already in operation, and it is desired to find the excitation coefficients from measurements upon the far-zone pattern alone. Evidently the processes required are exactly the same as in the design problem with the exception that the function $u=\log|S|$ is obtained as a result of experimental evidence rather than by arbitrary specification. The number of roots involved is known from the beginning, since it is one less than the number of elements and convergence to a solution, especially by the analog method, should be rapid and exact. Lack of uniqueness which was of benefit in the design problem (because it allowed a choice among several possible arrays) will be a stumbling block in the evaluation problem, since it is not known which of the possible sequences of excitation coefficients is actually present in the array being studied. If the number of solutions is small, the correct one may be found by eliminating those which are obviously unreasonable, or by making some additional measurements directly on the array, if this is possible. If the number of solutions is large, it is probably impractical to attempt to find the correct one. If the array is such that all the roots of the space factor lie upon the unit circle, the solution is unique and the evaluation problem is extremely simple. In this case $\log|S|$ need not even be measured at all points; the positions of the cones of silence are sufficient to establish the root values and hence the excitation coefficients of the array.

In the foregoing discussion we have assumed that the array to be evaluated is one for which at least one complete period of the space factor is visible and therefore determinable by physical measurements. Such is

[11] See, for example, Philip Franklin, *Differential Equations for Electrical Engineers* (John Wiley and Sons, Inc., New York, 1933), p. 275.

the case if the spacing is greater than or equal to a half-wavelength. If the spacing is less than a half-wavelength, only a part of one period of the space factor is available, and the case is more difficult. The fact that the the total number of roots is known will be a help, and it is conceivable that the skilled operator will be able to say much about the complete period of $\log|S|$ even though he begins by knowing only a part of it.

IX. CONCLUSIONS

1. The linear array design problem is essentially the problem of finding a polynomial whose magnitude, when evaluated upon the unit circle, fulfills given requirements.

2. The design problem can be solved by the use of an electrical analog device which continually plots the logarithm of the magnitude of the polynomial in question in response to the operator's selection of its roots. The operator adjusts the positions of these roots in the complex plane until the output of the device is as desired. The coefficients of the polynomial, which correspond to the relative currents in the elements of the linear array, are then found by elementary methods.

3. The solution to the array design problem is not, in general, unique.

4. What can be done by the electrical analog can also be done by an analytical method; however, certain steps are extremely laborious. It is felt that the analog method will be more satisfactory from an engineering point of view for several reasons. For example, it is better adapted to the form in which the problem is likely to be stated, and is capable of investigating minor changes with comparative ease.

5. Closely related to the design problem is the evaluation problem which begins with an experimental determination of the magnitude of the space factor of an existing array and yields as a result the excitation coefficients of that array. This is essentially the same as the design problem as far as the mechanics of solution are concerned, and the work may be performed on the analog. Lack of uniqueness in the solution, however, will limit the utility of this technique.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Theory of the Circular Diffraction Antenna*

HAROLD LEVINE AND CHARLES H. PAPAS
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received August 21, 1950)

The circular diffraction antenna consists of a coaxial wave guide fitted with an infinite-plane conducting baffle, and open to free space. An equivalent circuit description, appropriate to principal-mode propagation in the coaxial region, is investigated theoretically. Variational expressions for the circuit parameters are derived, and used for accurate numerical evaluation.

I. INTRODUCTION

THE circular diffraction antenna (Fig. 1) is a simple system amenable to detailed theoretical analysis, and offers the possibility of close experimental comparison. For convenience in analysis, the antenna is assumed to operate with a single propagating mode (the principal mode) in the coaxial region; waves incident on the open end through the coaxial guide are excited by a remote periodic source. With a perfectly conducting structure, part of the incident energy flux on the open end is returned to the source via reflected waves of the same type, and the remainder appears in free-space waves. A useful description of the principal-mode behavior is provided by an equivalent circuit, in which the stored energy of nonpropagating modes near the open end and the energy lost by radiation are accounted for by lumped circuit elements. This paper is concerned with variational methods for calculating the circuit parameters. Questions which relate primarily to the radiation aspects, such as its angular distribution, are not considered.

As the principal mode of the coaxial wave guide has no angular variation in any plane normal to its propagation direction, the geometrical structure of the antenna ensures that this symmetry is characteristic of the entire electromagnetic field. A further consequence of the geometrical symmetry is that the only nonvanishing electromagnetic field components in cylindrical coordinates are H_ϕ, E_ρ, E_z, the last being absent for the principal *TEM* mode. Since the Maxwell equations show that the steady-state electric field components are obtained from H_ϕ by differentiation, the latter serves as a basic scalar function.

The construction of field components, in accordance with the prescribed excitation and boundary condition on the conducting surfaces, is facilitated by dealing separately with coaxial and external regions. A continu-

* Part of the research reported in this document was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR), the Signal Corps of the U. S. Army, and the U. S. Air Force, under ONR Contract N5ori-76, T. O. 1.

FIG. 1. Coaxial line with baffle.

ous solution for all space is contingent on equality of the respective tangential components H_ϕ, E_ρ in the open end of the coaxial guide (called the aperture), with continuity of the normal component E_z then ensured. This schedule of aperture boundary conditions can be fitted into the formal development in different ways. The scalar function H_ϕ at any point in either region can be expressed in terms of integrals involving its z-derivative (proportional to E_ρ) in the aperture, the latter function having a null value on the conducting baffle. All continuity requirements are satisfied by equating the two functions H_ϕ at each point of the aperture, with the resulting integral equation to determine the aperture field E_ρ, and thence the electromagnetic field everywhere. Alternatively, H_ϕ in the coaxial region is given by integrals containing its aperture values, whereas the function outside is determined by these values together with the distribution on the baffle. If the z-derivative of the respective functions H_ϕ are set equal in the aperture, the field component E_ρ is continuous there; moreover, the z-derivative of H_ϕ, and hence E_ρ, is required to vanish on the baffle. Thus a pair of dual integral equations for the distribution of H_ϕ in the aperture plane is obtained.

For an engineering description, a principal-mode current and voltage in the coaxial guide are introduced. The current and voltage, in any cross section of the guide, are given by the line integrals, respectively, of the principal-mode part of H_ϕ around the inner conductor and the corresponding part of E_ρ between inner and outer conductors; the functions thus defined satisfy uniform transmission-line equations. An equivalent circuit representation, which employs the open end of the coaxial guide as a reference plane, involves a terminal admittance for the transmission line, specified by the current-voltage ratio in the aperture. Real and imaginary parts of the admittance are measures of the radiated and stored energy, respectively.

By combining the transmission-line quantities with the integral equation formulations, a pair of variational expressions for the terminal admittance (relative to the characteristic admittance of the line) are derived.[1] These have the property of being stationary with respect to variations about the aperture field E_ρ or the field H_ϕ in the aperture plane, and can be exhibited in useful homogeneous forms. Application of the variational expressions is given, with rather satisfactory experimental comparison.

II. FIRST INTEGRAL EQUATION FORMULATION

In accordance with the preceding section, the electromagnetic field can be derived from a single scalar function H_ϕ which is independent of the angular coordinate ϕ. For harmonic time dependence $\exp(-i\omega t)$ the differential equation which governs the spatial behavior of H_ϕ is

$$\left(\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}+\frac{\partial^2}{\partial z^2}-\frac{1}{\rho^2}+k^2\right)H_\phi(\rho,z)=0, \quad (2.1)$$

where $k^2=\omega^2\epsilon\mu=2\pi/\lambda$, ω=angular frequency, λ=free-space wavelength, ϵ and μ denote respectively the dielectric constant and permeability of free space. The electric-field components are derived from H_ϕ according to the Maxwell equations

$$i\omega\epsilon E_\rho(\rho,z)=\frac{\partial}{\partial z}H_\phi(\rho,z), \quad (2.2)$$

$$-i\omega\epsilon E_z(\rho,z)=\left(\frac{\partial}{\partial\rho}+\frac{1}{\rho}\right)H_\phi(\rho,z). \quad (2.3)$$

In the coaxial region, $a\leq\rho\leq b$, $z\leq 0$, H_ϕ must satisfy the boundary condition

$$\left(\frac{\partial}{\partial\rho}+\frac{1}{\rho}\right)H_\phi(\rho,z)=0 \quad \text{at } \rho=a,\ b, \quad (2.4)$$

which, according to (2.3), makes E_z vanish at the surfaces of the inner and outer conductors. A general expansion for H_ϕ is

$$H_\phi^{(-)}(\rho,z)=\alpha\frac{e^{ikz}}{\rho}+\beta\frac{e^{-ikz}}{\rho}+\sum_{n=1}^{\infty}A_nR_n(\rho)\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z], \quad a\leq\rho\leq b,\ z\leq 0, \quad (2.5)$$

where the first and second terms represent the incident and reflected TEM modes of amplitudes α and β respectively and the infinite series represents the higher order modes. The eigenfunction $R_n(\rho)$ and the eigenvalue λ_n are defined by

$$\left(\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}-\frac{1}{\rho^2}+\lambda_n^2\right)R_n(\rho)=0, \quad (2.6)$$

$$\left(\frac{\partial}{\partial\rho}+\frac{1}{\rho}\right)R_n(\rho)=0 \quad \text{at } \rho=a,\ b. \quad (2.7)$$

[1] For some of the original wave-guide applications of the variational method, see *Notes on Lectures by J. Schwinger* by David S. Saxon, Massachusetts Institute of Technology, February 1945.

The eigenvalues are roots of the transcendental equation

$$J_0(\lambda_n a)N_0(\lambda_n b)=N_0(\lambda_n a)J_0(\lambda_n b), \tag{2.8}$$

where $J_0(x)$ and $N_0(x)$ are the Bessel and Neumann functions of order zero, and a, b are the respective radii of the inner and outer conductors. Since $\lambda_{n+1}>\lambda_n$ and k is chosen such that $k<\lambda_1$, $(\lambda_n^2-k^2)^{\frac{1}{2}}$ is real and positive.[2] Consequently, all the higher order modes of (2.5) are exponentially damped with increasingly negative values of z. A corollary of (2.6) and (2.7) is the orthogonality of the eigenfunctions with one another and with $1/\rho$. By an appropriate normalization

$$\int_a^b R_n(\rho)R_m(\rho)\rho d\rho=\delta_{nm}, \tag{2.9}$$

$$\int_a^b R_n(\rho)d\rho=0, \tag{2.10}$$

where $\delta_{nn}=1$, $\delta_{nm}=0$. With the aid of (2.9) and (2.10) the coefficients A_n of (2.5) are expressed in terms of the transverse component of the electric-field intensity in the aperture. Thus applying (2) to (5) we obtain

$$i\omega\epsilon E_\rho^{(-)}(\rho, z)=\frac{\partial}{\partial z}H_\phi^{(-)}(\rho, z)=ik\alpha\frac{e^{ikz}}{\rho}-ik\beta\frac{e^{-ikz}}{\rho}$$
$$+\sum_{n=1}^{\infty}A_nR_n(\rho)(\lambda_n^2-k^2)^{\frac{1}{2}}$$
$$\times\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z], \quad a\leq\rho\leq b, \quad z\leq 0, \tag{2.11}$$

which, upon setting $z=0$, multiplying by $\rho R_n(\rho)$, and integrating from $\rho=a$ to $\rho=b$, yields

$$A_n=\frac{i\omega\epsilon}{(\lambda_n^2-k^2)^{\frac{1}{2}}}\int_a^b \mathcal{E}(\rho)R_n(\rho)\rho d\rho, \tag{2.12}$$

where $\mathcal{E}(\rho)\equiv E_\rho(\rho, 0)$. Substituting these A_n's into (2.5) we have

$$H_\phi^{(-)}(\rho, z)=\alpha\frac{e^{ikz}}{\rho}+\beta\frac{e^{-ikz}}{\rho}+i\omega\epsilon\int_a^b \mathcal{E}(\rho')\rho' d\rho'$$
$$\times\sum_{n=1}^{\infty}\frac{R_n(\rho)R_n(\rho')}{(\lambda_n^2-k^2)^{\frac{1}{2}}}\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z]$$
$$a\leq\rho\leq b, \quad z\leq 0, \tag{2.13}$$

A direct integration of (2.11) for $z=0$ provides the relation

$$\int_a^b \mathcal{E}(\rho)d\rho=(\alpha-\beta)\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\log\frac{b}{a}. \tag{2.14}$$

Our next task is to find an expression for the azimuthal component $H_\phi^{(+)}(\rho, z)$ of the magnetic-field intensity in the half-space $z\geq 0$. For this purpose we need a Green's function $G^{(+)}(\rho, z, \rho', z')$ which satisfies the inhomogeneous equation

$$\left(\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}+\frac{\partial^2}{\partial z^2}-\frac{1}{\rho^2}+k^2\right)G^{(+)}(\rho, z, \rho', z')$$
$$=-\delta(z-z')\frac{\delta(\rho-\rho')}{\rho}, \quad z\geq 0, \tag{2.15}$$

and the boundary condition

$$\frac{\partial}{\partial z}G^{(+)}(\rho, 0, \rho', z')=0. \tag{2.16}$$

We construct $G^{(+)}(\rho, \rho', z, z')$ by combining two solutions, one of which satisfies (2.15) for all z while the other satisfies (2.15) with the sign of z' changed, such that their sum is an even function of z [thereby satisfying (2.16)]. Thus

$$G^{(+)}(\rho, z, \rho', z')=G(\rho, \rho', z-z')+G(\rho, \rho', z+z'), \quad z\geq 0, \tag{2.17}$$

and $G(\rho, \rho', z\mp z')$ must satisfy the radiation condition and the equation

$$\left(\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}+\frac{\partial^2}{\partial z^2}-\frac{1}{\rho^2}+k^2\right)G(\rho, \rho', z\mp z')$$
$$=-\delta(z\mp z')\frac{\delta(\rho-\rho')}{\rho}, \quad -\infty\leq z\leq\infty, \tag{2.18}$$

where the upper and lower signs correspond to a source at $z=z'$, $\rho=\rho'$ and its image at $z=-z'$, $\rho=\rho'$ respectively. To determine $G(\rho, \rho', z-z')$ we take the Fourier transform[3] of (2.18),

$$\left[\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}-\zeta^2-\frac{1}{\rho^2}+k^2\right]G(\rho, \rho', \zeta)=\frac{-\delta(\rho-\rho')}{\rho}, \tag{2.19}$$

where ζ is a complex variable and

$$G(\rho, \rho', \zeta)=\int_{-\infty}^{\infty}e^{-i\zeta z}G(\rho, \rho', z)dz.$$

The solution for (2.19) representing cylindrical wave propagation outwards along ρ is

$$G(\rho, \rho', \zeta)=\frac{\pi i}{2}H_1^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}\rho_>]J_1[(k^2-\zeta^2)^{\frac{1}{2}}\rho_<], \tag{2.20}$$

where $\rho_>$, $\rho_<$ respectively designate the larger and smaller of ρ, ρ', $J_1(x)$ is the first-order Bessel function, and $H_1^{(1)}(x)=J_1(x)+iN_1(x)$ is the first-order Hankel function of the first kind. We specify the phases of the radicals in (2.20) as follows: $\arg(k^2-\zeta^2)^{\frac{1}{2}}=0, \pi/2$ for

[2] M. Kline, J. Math. Phys. **27**, 37–48, No. 1 (1949); H. B. Dwight, J. Math. Phys. **27**, 84–89, No. 1 (1949).

[3] H. Levine and J. Schwinger, Phys. Rev. **73**, 383–406 (1948).

$|\zeta|<k$, $|\zeta|>k$ respectively, to insure the proper sense of propagation. Applying the Fourier inversion theorem to (2.20) we obtain (Appendix A)

$$G(\rho, \rho', z-z') = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\rho, \rho', \zeta)e^{i\zeta(z-z')}d\zeta$$

$$= \frac{i}{4}\int_{-\infty}^{\infty} H_1^{(1)}[(k^2-\zeta^2)^{\frac{1}{2}}\rho_>]J_1[(k^2-\zeta^2)^{\frac{1}{2}}\rho_<]e^{i\zeta(z-z')}d\zeta$$

$$= \frac{1}{4\pi}\int_0^{2\pi} d\phi\cos\phi\frac{\exp\{ik[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z-z')^2]^{\frac{1}{2}}\}}{[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z-z')^2]^{\frac{1}{2}}}, \tag{2.21}$$

where the ζ-integration is along the real axis of the complex ζ-plane with an upward indentation at $\zeta=-k$ and a downward one at $\zeta=+k$ (Fig. 2). And by a similar procedure,

$$G(\rho, \rho', z+z') = \frac{1}{4\pi}\int_0^{2\pi} d\phi\cos\phi\frac{\exp\{ik[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z+z')^2]^{\frac{1}{2}}\}}{[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z+z')^2]^{\frac{1}{2}}}. \tag{2.22}$$

Hence, from (2.17), (2.21), and (2.22) we obtain the desired Green's function,

$$G^{(+)}(\rho, \rho', z, z') = \frac{1}{4\pi}\int_0^{2\pi} d\phi\cos\phi\left[\frac{\exp\{ik[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z-z')^2]^{\frac{1}{2}}\}}{[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z-z')^2]^{\frac{1}{2}}}\right.$$

$$\left.+\frac{\exp\{ik[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z+z')^2]^{\frac{1}{2}}\}}{[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z+z')^2]^{\frac{1}{2}}}\right]. \tag{2.23}$$

Now we apply a modified form of Green's second identity to $H_\phi^{(+)}(\rho, z)$ and $G^{(+)}(\rho, \rho', z, z')$ in the volume V bounded by the plane S of the baffle and the hemisphere S_∞ with $r\to\infty$, (Fig. 3). Thus

$$\int_V \{G^{(+)}(\rho, \rho', z, z')[\nabla'^2-1/\rho'^2+k^2]H_\phi^{(+)}(\rho', z') - H_\phi^{(+)}(\rho', z')[\nabla'^2-1/\rho'^2+k^2]G^{(+)}(\rho, \rho', z, z')\}d\tau'$$

$$= \int_{S+S_\infty}\{G^{(+)}(\rho, \rho', z, z')n\cdot\nabla'H_\phi^{(+)}(\rho', z') - H_\phi^{(+)}(\rho', z')n\cdot\nabla'G^{(+)}(\rho, \rho', z, z')\}d\sigma', \tag{2.24}$$

where n is a unit vector normal to the bounding surface and directed outward from the region. The integration over S_∞ disappears and, hence, the surface integral reduces to an integration over the plane $z=0$. By virtue of (2.1) and (2.15) the volume integration yields $H_\phi^{(+)}(\rho, z)$; recalling (2.2), (2.16), (2.23), and the boundary condition $E_\rho(\rho, 0)=0$, $b\le\rho\le a$, (2.24) becomes

$$H_\phi^{(+)}(\rho, z) = -\int_S G^{(+)}(\rho, \rho', z, 0)\frac{\partial}{\partial z'}H_\phi^{(+)}(\rho', 0)\rho'd\rho'$$

$$= \frac{-i\omega\epsilon}{2\pi}\int_a^b \mathcal{E}(\rho')\rho'd\rho'\int_0^{2\pi} d\phi\cos\phi\frac{\exp[ik(\rho^2+\rho'^2-2\rho\rho'\cos\phi+z^2)^{\frac{1}{2}}]}{(\rho^2+\rho'^2-2\rho\rho'\cos\phi+z^2)^{\frac{1}{2}}}, \quad z\ge 0. \tag{2.25}$$

Up to this point we have obtained expressions (2.13) and (2.25) for $H_\phi(\rho, z)$ valid in the coaxial region $a\le\rho\le b$, $z\le 0$ and the half-space $z\ge 0$ respectively. To proceed further we introduce the quantity of physical interest, i.e., the terminal admittance, in terms of the electric- and magnetic-field intensities.

As a measure of the coaxial-line current we take

$$I(z) = 2\pi(\alpha e^{ikz}+\beta e^{-ikz}), \tag{2.26}$$

which upon comparison with (2.13) is seen to be the integral of the dominant part of $H_\phi^{(-)}(\rho, z)$ taken along a circle of radius ρ surrounding the center conductor. And as a measure of the voltage across the coaxial line we adopt the line integral of $E_\rho^{(-)}(\rho, z)$ from $\rho=a$ to $\rho=b$,

$$V(z) = \int_a^b E_\rho^{(-)}(\rho, z)d\rho$$

$$= \left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}(\alpha e^{ikz}-\beta e^{-ikz})\log\frac{b}{a}, \tag{2.27}$$

FIG. 2. Path of integration in complex ζ-plane.

which follows from (2.11) and the orthogonality relation (2.10). In particular

$$V(0)=\int_a^b \mathcal{E}(\rho)d\rho=\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}(\alpha-\beta)\log\frac{b}{a}, \quad (2.28)$$

and

$$I(0)=2\pi(\alpha+\beta). \quad (2.29)$$

The current and voltage as defined by (2.26) and (2.27) satisfy the uniform transmission-line equations,

$$(\partial/\partial z)I(z)=ikY_0V(z), \quad (2.30)$$

$$(\partial/\partial z)V(z)=ikZ_0I(z), \quad (2.31)$$

where $Y_0(=1/Z_0)$ is the characteristic admittance and is given by

$$Y_0=\frac{2\pi}{(\mu/\epsilon)^{\frac{1}{2}}\log(b/a)}. \quad (2.32)$$

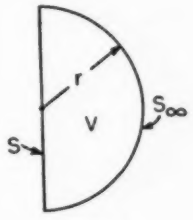


FIG. 3. Region of integration. S=plane of baffle. S_∞=hemisphere with $r=\infty$. V=volume enclosed by $S_\infty+S$.

The admittance $Y(z)$ at any cross section is given by the current-voltage ratio $Y(z)=I(z)/V(z)$. Taking the open end of the coaxial line as a reference plane we obtain an equivalent-circuit representation consisting of a lumped admittance $Y(0)$ terminating the transmission line. From (2.28), (2.29), and (2.32) we see that the terminal admittance is given by

$$Y(0)=\frac{I(0)}{V(0)}=Y_0\frac{\alpha+\beta}{\alpha-\beta}. \quad (2.33)$$

A knowledge of $Y(0)$ offers a complete description of the dominant-mode behavior.

Returning to the main thread of the discussion we see that $H_\phi^{(-)}(\rho, z)$ as given by (2.13) can be rewritten with the aid of (2.26) as

$$H_\phi^{(-)}(\rho, z)=\frac{I(z)}{2\pi\rho}+i\omega\epsilon\int_a^b \mathcal{E}(\rho')\rho'd\rho'\sum_{n=1}^{\infty}\frac{R_n(\rho)R_n(\rho')}{(\lambda_n^2-k^2)^{\frac{1}{2}}}\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z], \quad z\leq 0. \quad (2.34)$$

Across the aperture $H_\phi(\rho, z)$ must be continuous, i.e., $H_\phi^{(-)}(\rho, 0)=H_\phi^{(+)}(\rho, 0)$. Consequently setting $z=0$ in (2.25) and (2.34) we obtain the desired integral equation

$$\frac{I(0)}{2\pi\rho}+i\omega\epsilon\int_a^b \mathcal{E}(\rho')\rho'd\rho'\sum_{n=1}^{\infty}\frac{R_n(\rho)R_n(\rho')}{(\lambda_n^2-k^2)^{\frac{1}{2}}}$$

$$=\frac{-i\omega\epsilon}{2\pi}\int_a^b \mathcal{E}(\rho')\rho'd\rho'\int_0^{2\pi}d\phi\cos\phi\frac{\exp[ik(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}]}{(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}}, \quad a\leq\rho\leq b. \quad (2.35)$$

III. SECOND INTEGRAL EQUATION FORMULATION

Using an alternative existence theorem, which allows the representation of the electromagnetic field within a region in terms of the tangential magnetic field on the bounding surface, we now deduce a pair of dual equations for the distribution of the magnetic-field intensity $\mathcal{H}(\rho)$ in the aperture plane.

We recall from (2.5) that

$$H_\phi^{(-)}(\rho, z)=\alpha\frac{e^{ikz}}{\rho}+\beta\frac{e^{-ikz}}{\rho}+\sum_{n=1}^{\infty}A_nR_n(\rho)\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z], \quad a\leq\rho\leq b, \quad z\leq 0. \quad (2.5)$$

Multiplying by $\rho R_n(\rho)$, then integrating from $\rho=a$ to $\rho=b$, we obtain by (2.9), (2.10) for $z=0$

$$A_m=\int_a^b \mathcal{H}(\rho)R_m(\rho)\rho d\rho, \quad (3.1)$$

where $\mathcal{H}(\rho)\equiv H_\phi^{(-)}(\rho, 0)$. Similarly by a direct integration of (2.5)

$$\int_a^b \mathcal{H}(\rho)d\rho=(\alpha+\beta)\log\frac{b}{a}=\frac{I(0)}{2\pi}\log\frac{b}{a}, \quad (3.2)$$

where (2.29) was used to establish the latter equality. Substituting (3.1) into (2.5) we find

$$H_\phi^{(-)}(\rho, z)=\alpha\frac{e^{ikz}}{\rho}+\beta\frac{e^{-ikz}}{\rho}+\int_a^b \mathfrak{IC}(\rho')\rho' d\rho' \sum_{n=1}^{\infty} R_n(\rho)R_n(\rho')\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z], \quad a\leq\rho\leq b, \quad z\leq 0,$$

which, when differentiated according to (2.2) yields

$$E_\rho^{(-)}(\rho, z)=\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\left(\alpha\frac{e^{ikz}}{\rho}-\beta\frac{e^{-ikz}}{\rho}\right)$$

$$+\frac{1}{i\omega\epsilon}\int_a^b \mathfrak{IC}(\rho')\rho' d\rho' \sum_{n=1}^{\infty} R_n(\rho)R_n(\rho')(\lambda_n^2-k^2)^{\frac{1}{2}}\exp[(\lambda_n^2-k^2)^{\frac{1}{2}}z], \quad a\leq\rho\leq b, \quad z\leq 0. \quad (3.3)$$

To derive an expression for the transverse component $E_\rho^{(+)}(\rho, z)$ of the electric-field intensity for the half-plane $z\geq 0$, we construct a Green's function $\Gamma(\rho, \rho', z, z')$ which satisfies the same differential equation (2.15) as before, but with a different boundary condition,

$$\Gamma(\rho, \rho', 0, z')=0. \quad (3.4)$$

Thus

$$\Gamma(\rho, \rho', z, z')=G(\rho, \rho', z-z')-G(\rho, \rho', z+z')$$

$$=\frac{1}{4\pi}\int_0^{2\pi} d\phi\cos\phi\left[\frac{\exp\{ik[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z-z')^2]^{\frac{1}{2}}\}}{[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z-z')^2]^{\frac{1}{2}}}\right.$$

$$\left.-\frac{\exp\{ik[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z+z')^2]^{\frac{1}{2}}\}}{[\rho^2+\rho'^2-2\rho\rho'\cos\phi+(z+z')^2]^{\frac{1}{2}}}\right]. \quad (3.5)$$

From Green's identity (2.24), with $G^{(+)}(\rho, \rho', z, z')$ replaced by $\Gamma(\rho, \rho', z, z')$, it follows that

$$H_\phi^{(+)}(\rho, z)=\int_0^\infty H_\phi^{(+)}(\rho', 0)\frac{\partial}{\partial z'}\Gamma(\rho, \rho', z, 0)\rho' d\rho', \quad z\geq 0, \quad (3.6)$$

from which we get, with the aid of (2.2),

$$E_\rho^{(+)}(\rho, z)=\frac{1}{i\omega\epsilon}\int_0^\infty \mathfrak{IC}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', z, 0)\rho' d\rho', \quad z\geq 0, \quad (3.7)$$

where $\mathfrak{IC}(\rho)\equiv H_\phi^{(+)}(\rho, 0)$.

In the aperture $E_\rho^{(-)}(\rho, 0)=E_\rho^{(+)}(\rho, 0)$. Hence by (3.3) and (3.7)

$$\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\left(\frac{\alpha}{\rho}-\frac{\beta}{\rho}\right)+\frac{1}{i\omega\epsilon}\int_a^b \mathfrak{IC}(\rho')\rho' d\rho' \sum_{n=1}^{\infty} R_n(\rho)R_n(\rho')(\lambda_n^2-k^2)^{\frac{1}{2}}$$

$$=\frac{1}{i\omega\epsilon}\int_0^\infty \mathfrak{IC}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', 0, 0)\rho' d\rho', \quad a\leq\rho\leq b, \quad (3.8)$$

and on the remainder of the plane $E_\rho^{(+)}(\rho, 0)$ must vanish:

$$0=\int_0^\infty \mathfrak{IC}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', 0, 0)\rho' d\rho', \quad a\geq\rho\geq b. \quad (3.9)$$

Multiplying (3.8) and (3.9) by $\rho\mathfrak{IC}(\rho)$ and integrating,

$$\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}(\alpha-\beta)\int_a^b \mathfrak{IC}(\rho)d\rho+\frac{1}{i\omega\epsilon}\sum_{n=1}^{\infty}(\lambda_n^2-k^2)^{\frac{1}{2}}\left[\int_a^b \mathfrak{IC}(\rho)R_n(\rho)\rho d\rho\right]^2$$

$$=\frac{1}{i\omega\epsilon}\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)\int_0^\infty \rho' d\rho'\mathfrak{IC}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', 0, 0). \quad (3.10)$$

We rewrite (3.10) in a form involving the terminal voltage $V(0)$ given by (2.28):

$$\frac{V(0)}{\log(b/a)}\int_a^b \mathcal{K}(\rho)d\rho+\frac{1}{i\omega\epsilon}\sum_{n=1}^{\infty}(\lambda_n^2-k^2)^{\frac{1}{2}}\left[\int_a^b \mathcal{K}(\rho)R_n(\rho)\rho d\rho\right]^2$$

$$=\frac{1}{i\omega\epsilon}\int_0^\infty \rho d\rho \mathcal{K}(\rho)\int_0^\infty \rho' d\rho' \mathcal{K}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho,\rho',0,0). \quad (3.11)$$

IV. VARIATIONAL PRINCIPLES

Utilizing the first integral equation, we construct a variational principle for $Y(0)$ by first multiplying (2.35) by $\rho\mathcal{E}(\rho)$, then integrating from $\rho=a$ to $\rho=b$, and finally dividing the resultant equation by $[\int_a^b \mathcal{E}(\rho)d\rho]^2$. Thus, recalling that $V(0)=\int_a^b \mathcal{E}(\rho)d\rho$ (2.28) and $Y(0)=I(0)/V(0)$ (2.33), we obtain the desired representation,

$$\frac{Y(0)}{Y_0}=\frac{-ik\log\frac{b}{a}}{\left[\int_a^b \mathcal{E}(\rho)d\rho\right]^2}\left\{\sum_{n=1}^{\infty}\frac{\left[\int_a^b \mathcal{E}(\rho)R_n(\rho)\rho d\rho\right]^2}{(\lambda_n^2-k^2)^{\frac{1}{2}}}\right.$$

$$\left.+\frac{1}{2\pi}\int_a^b \rho d\rho\int_a^b \rho' d\rho'\int_0^{2\pi} d\phi\cos\phi\mathcal{E}(\rho)\mathcal{E}(\rho')\frac{\exp[ik(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}]}{(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}}\right\}, \quad (4.1)$$

for $Y(0)/Y_0$ in terms of the aperture electric field $\mathcal{E}(\rho)$, which is homogeneous in $\mathcal{E}(\rho)$ and stationary with respect to the first variation of $\mathcal{E}(\rho)$ about its correct value determined by (2.35).

Another variational principle is deduced by dividing (3.11) by $[\int_a^b \mathcal{K}(\rho)d\rho]^2$. Thus, since $2\pi\int_a^b \mathcal{K}(\rho)d\rho=I(0)\times\log b/a$, $Y(0)=I(0)/V(0)$,

$$\frac{Y_0}{Y(0)}=\frac{\log\frac{b}{a}}{ik\left[\int_a^b \mathcal{K}(\rho)d\rho\right]^2}\left\{\int_0^\infty \rho d\rho\mathcal{K}(\rho)\int_0^\infty \rho' d\rho'\mathcal{K}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho,\rho',0,0)\right.$$

$$\left.-\sum_{n=1}^{\infty}(\lambda_n^2-k^2)^{\frac{1}{2}}\left[\int_a^b \mathcal{K}(\rho)R_n(\rho)\rho d\rho\right]^2\right\}, \quad (4.2)$$

which is homogeneous in $\mathcal{K}(\rho)$ and stationary with respect to first variations about its correct value determined by (3.11).

In both variational principles the stationary quantity, $Y(0)=G(0)-iB(0)$ where $G(0)$ and $B(0)$ are the terminating conductance and susceptance respectively, is complex. If we substitute the true $\mathcal{E}(\rho)$ in (4.1) and the true $\mathcal{K}(\rho)$ in (4.2), obtained from (2.35) and (3.11), respectively, both variational principles will yield the same true value for $Y(0)$. However, the two values of $Y(0)$ resulting from approximate $\mathcal{E}(\rho)$ and $\mathcal{K}(\rho)$ will differ, but both will be close to the true $Y(0)$ if the trial functions are chosen appropriately. The errors in $Y(0)$ are of the order of magnitude of the squares of the errors in $\mathcal{E}(\rho)$ and $\mathcal{K}(\rho)$.

If the stationary quantity had not been complex, one variational principle would yield a value above the true one while the other would yield a value below it. Since $Y(0)$ is complex no such limits can be found.

V. APPLICATION OF THE VARIATIONAL PRINCIPLES

We evaluate $Y(0)$ using first (4.1) involving $\mathcal{E}(\rho)$ and then (4.2) involving $\mathcal{K}(\rho)$.

Assuming

$$\mathcal{E}(\rho)=\sum_{n=0}^{\infty}a_nR_n(\rho), \quad (5.1)$$

where $R_0(\rho)=1/\rho$, we calculate the integrals of (4.1) in terms of a_n. Recalling the orthogonality properties (2.9),

(2.10) of $R_n(\rho)$, we directly get

$$\left[\int_a^b \mathcal{E}(\rho)d\rho\right]^2=\left[\sum_{n=0}^{\infty} a_n\int_a^b R_n(\rho)d\rho\right]^2=a_0^2\left(\log\frac{b}{a}\right)^2, \tag{5.2}$$

and

$$\left[\int_a^b \mathcal{E}(\rho)R_n(\rho)\rho d\rho\right]^2=\left[\sum_{m=0}^{\infty} a_m\int_a^b R_m(\rho)R_n(\rho)\rho d\rho\right]^2=\left[\sum_{m=0}^{\infty} a_m\delta_{mn}\right]^2=a_n^2. \tag{5.3}$$

The remaining integral of (4.1) is cast into a more convenient form by the relation (Appendix A)

$$\frac{1}{2\pi}\int_0^{2\pi} d\phi\cos\phi\frac{\exp[ik(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}]}{(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}}=\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}J_1(\zeta\rho)J_1(\zeta\rho'), \tag{5.4}$$

where the ζ-integration is along the real axis of the complex ζ-plane with a downward indentation at $\zeta=+k$ and $\arg(\zeta^2-k^2)^{\frac{1}{2}}=0$, $-\pi/2$ for $\zeta\geq k$, $\zeta\leq k$ respectively. Thus

$$\frac{1}{2\pi}\int_a^b \rho d\rho\int_a^b \rho' d\rho'\int_0^{2\pi} d\phi\cos\phi\mathcal{E}(\rho)\mathcal{E}(\rho')\frac{\exp[ik(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}]}{(\rho^2+\rho'^2-2\rho\rho'\cos\phi)^{\frac{1}{2}}}=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_n a_m c_{nm}, \tag{5.5}$$

where

$$c_{nm}=\int_a^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}\int_a^b \rho d\rho\int_0^b \rho' d\rho' J_1(\zeta\rho)J_1(\zeta\rho')R_n(\rho)R_m(\rho')=c_{mn}. \tag{5.6}$$

Carrying out the ρ and ρ' integrations of (5.6) (Appendix B),

$$c_{nm}=\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}f_n(\zeta)f_m(\zeta), \tag{5.7}$$

where

$$f_n(\zeta)=\int_a^b \rho d\rho R_n(\rho)J_1(\zeta\rho)$$

$$=\frac{2}{\pi}\frac{N_n}{\lambda_n}\frac{1}{J_0(\lambda_n b)}\frac{\zeta}{\lambda_n^2-\zeta^2}[J_0(\zeta b)J_0(\lambda_n a)-J_0(\zeta a)J_0(\lambda_n b)], \tag{5.8}$$

and

$$f_0(\zeta)=\int_a^b \rho d\rho R_0(\rho)J_1(\zeta\rho)=\frac{J_0(\zeta a)-J_0(\zeta b)}{\zeta}, \tag{5.9}$$

and

$$\frac{1}{N_n}=\frac{\sqrt{2}}{\pi\lambda_n}\left[\frac{J_0^2(\lambda_n a)}{J_0^2(\lambda_n b)}-1\right]^{\frac{1}{2}}. \tag{5.10}$$

Inserting (5.2), (5.3), and (5.5) into (4.1) we get the following expression for $Y(0)$ in terms of the coefficients a_m (only the relative amplitudes a_m/a_0 are important for the calculation):

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\left\{\sum_{n=1}^{\infty}\frac{\bar{a}_n^2}{(\lambda_n^2-k^2)^{\frac{1}{2}}}+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\bar{a}_n\bar{a}_m c_{nm}+2\sum_{m=1}^{\infty}\bar{a}_m c_{0m}+c_{00}\right\}, \tag{5.11}$$

where $\bar{a}_m=a_m/a_0$.

Owing to the stationary character of $Y(0)$, on differentiating (5.11) with respect to $\bar{a}_n$, we obtain the following infinite set of simultaneous equations,

$$\frac{\bar{a}_n}{(\lambda_n^2-k^2)^{\frac{1}{2}}}+c_{0n}+\sum_{m=1}^{\infty}\bar{a}_m c_{nm}=0,\quad n=1, 2, 3, \cdots, \infty, \tag{5.12}$$

which, when multiplied by $\bar{a}_n$ and summed on n, yield

$$\sum_{n=1}^{\infty}\frac{\bar{a}_n^2}{(\lambda_n^2-k^2)^{\frac{1}{2}}}+\sum_{n=1}^{\infty}\bar{a}_n c_{0n}+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\bar{a}_n\bar{a}_m c_{nm}=0. \tag{5.13}$$

As a consequence of (5.13), (5.11) becomes

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\{c_{00}+\sum_{m=1}^{\infty}\bar{a}_m c_{0m}\}. \tag{5.14}$$

Now we proceed with the evaluation of $Y(0)$ using (4.2). In preparation for this, we deduce an appropriate representation for $\mathfrak{IC}(\rho)$. From (2.25) and (5.4),

$$H_\phi^{(+)}(\rho, z)=-i\omega\epsilon\int_0^\infty \frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}J_1(\zeta\rho)\int_a^b \rho' d\rho' J_1(\zeta\rho')\mathcal{E}(\rho'). \tag{5.15}$$

Inserting the expansion (5.1) into (5.15) we find for $\mathfrak{IC}(\rho)$ (apart from scale factors)

$$\mathfrak{IC}(\rho)=\sum_{m=0}^{\infty} a_m\int_0^\infty \frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}J_1(\zeta\rho)f_m(\zeta), \tag{5.16}$$

since $\mathfrak{IC}(\rho)\equiv H_\phi^{(+)}(\rho, 0)$. From (5.9) and (5.16)

$$\int_a^b \mathfrak{IC}(\rho)d\rho=\sum_{m=0}^{\infty} a_m\int_0^\infty \frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}f_m(\zeta)\int_a^b J_1(\zeta\rho)d\rho$$

$$=\sum_{m=0}^{\infty} a_m\int_0^\infty \frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}f_m(\zeta)f_0(\zeta)=\sum_{m=0}^{\infty} a_m c_{0m}, \tag{5.17}$$

from which it follows that

$$\left[\int_a^b \mathfrak{IC}(\rho)d\rho\right]^2=\left(\sum_{m=0}^{\infty} a_m c_{0m}\right)^2. \tag{5.18}$$

Since [Appendix A(13)]

$$\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', 0, 0)=-\int_0^\infty \zeta(\zeta^2-k^2)^{\frac{1}{2}}J_1(\zeta\rho)J_1(\zeta\rho')d\zeta, \tag{5.19}$$

we write

$$\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)\int_0^\infty \rho' d\rho'\mathfrak{IC}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', 0, 0)=-\int_0^\infty \zeta d\zeta(\zeta^2-k^2)^{\frac{1}{2}}\left[\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)J_1(\zeta\rho)\right]^2, \tag{5.20}$$

and proceed to evaluate $\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)J_1(\zeta\rho)$. Utilizing (5.16), we get

$$\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)J_1(\zeta\rho)=\int_0^\infty \rho d\rho J_1(\zeta\rho)\sum_{m=0}^{\infty} a_m\int_0^\infty \frac{\xi d\xi}{(\xi^2-k^2)^{\frac{1}{2}}}J_1(\xi\rho)f_m(\xi), \tag{5.21}$$

where ξ is a complex variable. Recognizing that

$$\int_0^\infty \rho d\rho J_1(\zeta\rho)J_1(\xi\rho)=\frac{\delta(\zeta-\xi)}{\zeta}, \tag{5.22}$$

(5.21) becomes

$$\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)J_1(\zeta\rho)=\sum_{m=0}^{\infty} a_m\frac{f_m(\zeta)}{(\zeta^2-k^2)^{\frac{1}{2}}}. \tag{5.23}$$

It follows from (5.7), (5.20), (5.23) that

$$\int_0^\infty \rho d\rho\mathfrak{IC}(\rho)\int_0^\infty \rho' d\rho'\mathfrak{IC}(\rho')\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho, \rho', 0, 0)$$

$$=-\int^\infty \zeta d\zeta(\zeta^2-k^2)^{\frac{1}{2}}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{a_n a_m}{\zeta^2-k^2}f_n(\zeta)f_m(\zeta)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_n a_m c_{nm}. \tag{5.24}$$

Next we evaluate $\int_a^b \mathcal{K}(\rho)R_n(\rho)\rho d\rho$. Again, from (5.16) we get

$$\int_a^b \mathcal{K}(\rho)R_n(\rho)\rho d\rho=\sum_{m=0}^{\infty} a_m\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}f_m(\zeta)\int_a^b \rho d\rho R_n(\rho)J_1(\zeta\rho), \tag{5.25}$$

which by (5.7) and (5.8) becomes

$$\int_a^b \mathcal{K}(\rho)R_n(\rho)\rho d\rho=\sum_{m=0}^{\infty} a_m c_{nm}. \tag{5.26}$$

With the aid of (5.18), (5.24), and (5.26) we get for (4.2)

$$\frac{Y_0}{Y(0)}=\frac{\log\frac{b}{a}}{-ik}\frac{\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_n a_m c_{nm}+\sum_{n=1}^{\infty}(\lambda_n^2-k^2)^{\frac{1}{2}}(\sum_{m=0}^{\infty} a_m c_{nm})^2}{(\sum_{m=0}^{\infty} a_m c_{0m})^2}. \tag{5.27}$$

This expression is stationary and upon differentiating it with respect to a_p we obtain

$$c_{0p}\sum_{m=0}^{\infty} a_m c_{0m}=\left(\frac{Y(0)}{Y_0}\frac{\log(b/a)}{-ik}\right)\left(\sum_{m=0}^{\infty} a_m c_{pm}+\sum_{n=1}^{\infty} c_{pn}(\lambda_n^2-k^2)^{\frac{1}{2}}\sum_{m=0}^{\infty} a_m c_{mn}\right),\quad p=0, 1, 2, 3, \cdots, \infty. \tag{5.28}$$

Then

$$c_{0p}\sum_{m=0}^{\infty} a_m c_{0m}=\left(\frac{Y(0)}{Y_0}\frac{\log(b/a)}{-ik}\right)\sum_{m=0}^{\infty} a_m d_{pm}, \tag{5.29}$$

where

$$d_{pm}=c_{pm}+\sum_{n=1}^{\infty} c_{pn}c_{mn}(\lambda_n^2-k^2)^{\frac{1}{2}}=d_{mp}. \tag{5.30}$$

We now define b_n by the relation

$$a_n=\frac{Y_0}{Y(0)}\frac{-ik}{\log(b/a)}b_n\sum_{m=0}^{\infty} a_m c_{0m}, \tag{5.31}$$

from which, after multiplication by c_{0n} and summation on n, it follows that

$$\sum_{n=0}^{\infty} a_n c_{0n}=\frac{Y_0}{Y(0)}\frac{-ik}{\log(b/a)}\sum_{n=0}^{\infty} b_n c_{0n}\sum_{n=0}^{\infty} a_m c_{0m}. \tag{5.32}$$

The sum $\sum_{n=0}^{\infty} a_n c_{0n}$ is common to both sides of (5.32); canceling it out we obtain

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\sum_{m=0}^{\infty} b_n c_{0n}. \tag{5.33}$$

Inserting (5.31) into (5.28) and using (5.33), we get the following equations for the b_n:

$$c_{0p}=\sum_{m=0}^{\infty} b_m d_{pm},\quad p=0, 1, 2, 3, \cdots, \infty. \tag{5.34}$$

The first and second variational principles have yielded (5.14) and (5.33) where the determination of the associated coefficients $\bar{a}_n$, b_n require the solution of two infinite sets of simultaneous equations (5.12) and (5.34). A rigorous solution is not attempted; instead we assume that $\bar{a}_n=b_n=0$ when $n\geq N$. The corresponding approximations to $Y(0)$ are given by

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\{c_{00}+\sum_{n=1}^{N}\bar{a}_n c_{0n}\}, \tag{5.35}$$

where

$$\frac{\bar{a}_n}{(\lambda_n^2-k^2)^{\frac{1}{2}}}+c_{0n}+\sum_{m=1}^{\infty}\bar{a}_m c_{nm}=0,\quad n=1, 2, 3, \cdots, N, \tag{5.36}$$

for the first variational principle, and

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\sum_{n=0}^{N} b_n c_{0n}, \tag{5.37}$$

where

$$\sum_{n=0}^{N} b_n d_{pn}=c_{0p}, \quad p=0, 1, 2, \cdots, N, \tag{5.38}$$

for the second.

The computational procedure indicated in (5.35)–(5.38) affords approximations to any desired degree. The simplest approximation (and yet a sufficiently accurate one, as we shall see upon comparison with experimental results) is obtained by neglecting all but the leading terms. Hence from (5.35) we obtain

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\int_0^\infty \frac{d\zeta}{\zeta(\zeta^2-k^2)^{\frac{1}{2}}}[J_0(\zeta a)-J_0(\zeta b)]^2, \tag{5.39}$$

and from (5.37)

$$\frac{Y(0)}{Y_0}=\frac{-ik}{\log(b/a)}\cdot\frac{\left(\int_0^\infty \frac{d\zeta}{\zeta(\zeta^2-k^2)^{\frac{1}{2}}}[J_0(\zeta a)-J_0(\zeta b)]^2\right)^2}{\int_0^\infty \frac{d\zeta}{\zeta(\zeta^2-k^2)^{\frac{1}{2}}}[J_0(\zeta a)-J_0(\zeta b)]^2+\sum_{n=1}^{\infty}(\lambda_n^2-k^2)^{\frac{1}{2}}\left\{\int_0^\infty \frac{d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}[J_0(\zeta a)-J_0(\zeta b)]f_n(\zeta)\right\}^2}, \tag{5.40}$$

where $f_n(\zeta)$ is given by (5.8).

If the approximation were improved by employing larger values of N in (5.35) and (5.37), the difference between the two resulting values for $Y(0)$ would progressively decrease and in the limit $N\to\infty$ both variational principles would yield the same true value. Comparing (5.39) with (5.40) it is seen that they differ by virtue of the term,

$$\sum_{n=1}^{\infty}(\lambda_n^2-k^2)^{\frac{1}{2}}\left\{\int_0^\infty \frac{d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}[J_0(\zeta a)-J_0(\zeta b)]f_n(\zeta)\right\}^2, \tag{5.41}$$

the magnitude of which offers a measure of accuracy for (5.39). That this term is small is placed in evidence by the fact that (5.41) converges at least as fast as $1/n^3$ (this directly follows from (5.41) and the asymptotic relation $\lambda_n\sim n\pi/(b-a)$ as $n\to\infty$) and by the close comparison between $Y(0)$ computed from (5.39) and experiment.

VI. EVALUATION OF THE CIRCUIT PARAMETERS†

The terminating conductance $G(0)$ and susceptance $B(0)$ are obtained from the real and imaginary parts of $Y(0)$ since $Y(0)=G(0)-iB(0)$. Thus, from (5.39)

$$\frac{Y(0)}{Y_0}=\frac{G(0)}{Y_0}-i\frac{B(0)}{Y_0}=\frac{-ik}{\log(b/a)}\int_0^\infty \frac{d\zeta}{\zeta(\zeta^2-k^2)^{\frac{1}{2}}}\times[J_0(\zeta a)-J_0(\zeta b)]^2. \tag{6.1}$$

The path of integration along the positive real axis of the complex ζ-plane with a downward indentation at $\zeta=k$ is considered as the sum of three separate integrations: one extending from $\zeta=0$ to $\zeta=k-\eta$, another along the semicircular indentation of radius η, and finally one from $\zeta=k+\eta$ to $\zeta=\infty$. As $\eta\to 0$ the integral along the semicircle vanishes. Therefore, (6.1) is the sum of two integrals, one extending from $\zeta=0$ to $\zeta=k$, the other from $\zeta=k$ to $\zeta=\infty$; using the transformations $\zeta=k\sin\theta$ and $\zeta=k\xi$ in the former and latter respectively, we obtain

$$\frac{G(0)}{Y_0}=\frac{1}{\log(b/a)}\int_0^{\pi/2}\frac{d\theta}{\sin\theta}\times[J_0(ka\sin\theta)-J_0(kb\sin\theta)]^2, \tag{6.2}$$

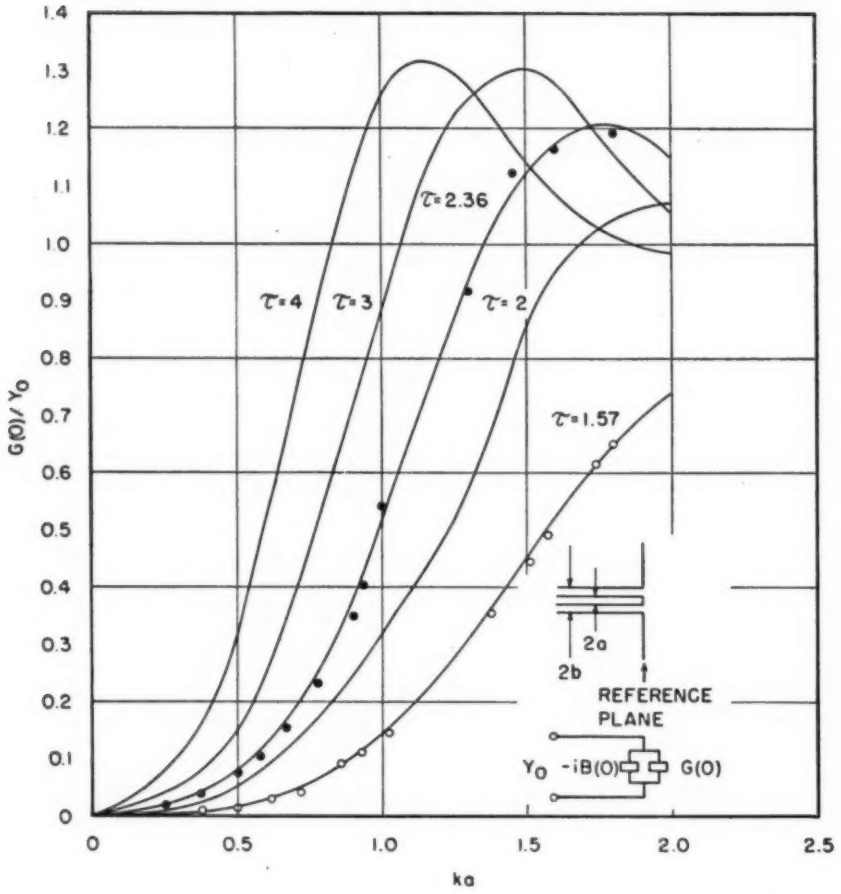
$$\frac{B(0)}{Y_0}=\frac{1}{\log(b/a)}\int_1^\infty \frac{d}{\xi(\xi^2-1)^{\frac{1}{2}}}[J_0(ka\xi)-J_0(kb\xi)]^2. \tag{6.3}$$

With the aid of the relations (Appendix C)

$$\int_z^\infty \frac{J_0(ka\xi)J_0(kb\xi)}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi=\frac{\pi}{2z}-\frac{1}{\pi z}\int_0^\pi Si[kz(a^2+b^2-2ab\cos\psi)^{\frac{1}{2}}]d\psi, \tag{6.4}$$

$$\int_z^\infty \frac{J_0^2(ka\xi)}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi=\frac{\pi}{2z}-\frac{1}{\pi z}\int_0^\pi Si(2kaz\sin\psi/2)d\psi, \tag{6.5}$$

† The analytical results of this section were obtained by one of us (H.L.) at the Radiation Laboratory, Massachusetts Institute of Technology, 1945 and are contained in the *Wave Guide Handbook*, Vol. 10 of the Laboratory Series (McGraw-Hill Book Company, Inc., New York).



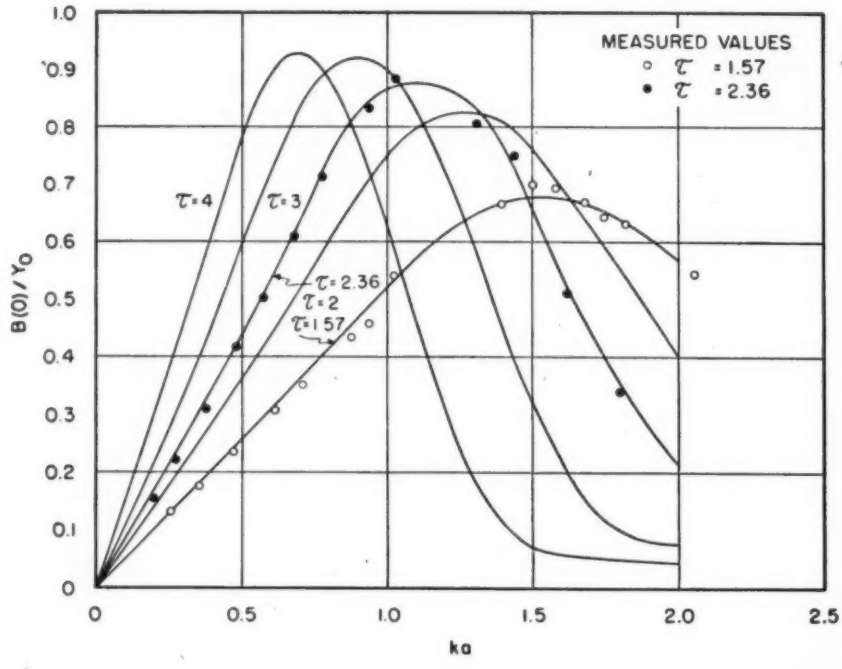


FIG. 4. Terminal conductance and susceptance (relative to characteristic admittance) as function of ka, $\tau=b/a$. $k=2\pi\div$free space wavelength.

where

$$Si(x)=\int_0^x \frac{\sin t}{t}dt$$

is the sine integral, (6.3) takes the more convenient form:

$$\frac{B(0)}{Y_0}=\frac{1}{\pi\log(b/a)}\times 2\int_0^\pi Si\{ka[1+(b/a)^2-2(b/a)\cos\psi]^{\frac{1}{2}}\}d\psi-\int_0^\pi Si(2ka\sin\psi/2)d\psi-\int_0^\pi Si(2kb\sin\psi/2)d\psi. \quad (6.6)$$

From (6.2) and (6.6) $G(0)/Y_0$ and $B(0)/Y_0$ have been computed for various values of ka with $\tau=b/a=1.57$, 2.0, 2.36, 3.0, and 4.0 and compared with experimental values[4] obtained for $\tau=1.57$ and 2.36. In Fig. 4 it is seen that the agreement between the calculated and measured values is satisfactory.

In the low frequency limit, i.e., ka, $kb\ll 1$, we obtain from (6.2) by retaining only the first two terms in the power series development of the Bessel functions

$$\frac{G(0)}{Y_0}\simeq\frac{2}{3}\frac{1}{\log\tau}\left(\frac{ka}{2}\right)^4(\tau^2-1)^2, \quad (6.7)$$

where $\tau=b/a$. Moreover, since $Si(x)\simeq x$ when $x\ll 1$, at low frequencies (6.6) reduces to

$$\frac{B(0)}{Y_0}=\frac{1}{\pi\log(b/a)}\times\left\{2ka\int_0^\pi\left[1+\left(\frac{b}{a}\right)^2-2\left(\frac{b}{a}\right)\cos\psi\right]^{\frac{1}{2}}d\psi-4(ka+kb)\right\}. \quad (6.8)$$

In terms of the complete elliptic integral of the second kind, $E(\alpha)$, defined by

$$E(\alpha)=\int_0^{\pi/2}(1-\alpha^2\sin^2\phi)^{\frac{1}{2}}d\phi, \quad (6.9)$$

(6.8) takes the form

$$\frac{B(0)}{Y_0}=\frac{4k(a+b)}{\pi\log(b/a)}\left\{E\left[\frac{2(ab)^{\frac{1}{2}}}{a+b}\right]-1\right\}=\frac{4ka(1+\tau)}{\pi\log\tau}\left\{E\left[\frac{2(\tau)^{\frac{1}{2}}}{\tau+1}\right]-1\right\}, \quad (6.10)$$

which is valid for ka, $kb\ll 1$. Furthermore, when $\alpha^2\equiv[2(\tau)^{\frac{1}{2}}/\tau+1]^2\simeq 1$ or $\tau\simeq 1$,

$$E(\alpha)\simeq 1+\frac{1}{2}\left[\log\frac{4}{(1-\alpha^2)^{\frac{1}{2}}}-\frac{1}{2}\right](1-\alpha^2). \quad (6.11)$$

Consequently, for ka, $kb\ll 1$, $\tau\simeq 1$ we obtain from (6.10) and (6.11)

$$\frac{B(0)}{Y_0}=\frac{2ka}{\pi\log\tau}\left[\log 4-\log\frac{\tau-1}{\tau+1}-\frac{1}{2}\right]. \quad (6.12)$$

Finally, we may note that in the limit ka, $kb\to\infty$

$$\frac{G(0)}{Y_0}-i\frac{B(0)}{Y_0}=\frac{-ik}{\log(b/a)}\int_0^\infty\frac{d\zeta}{\zeta(\zeta^2-k^2)^{\frac{1}{2}}}[J_0(\zeta a)-J_0(\zeta b)]^2\simeq\frac{1}{\log(b/a)}\int_0^\infty\frac{dx}{x}[J_0(x)-J_0(x\tau)]^2=1, \quad (6.13)$$

where the transformations $x=\zeta a$, $(\zeta^2-k^2)^{\frac{1}{2}}\to -ik$ are effected in going from the first integral to the second. Hence, in the high frequency limit $G(0)\simeq Y_0$ and $B(0)\simeq 0$ and the antenna is matched to free space for there is no reflection of the dominant mode wave back into the coaxial guide.

ACKNOWLEDGMENT

The authors wish to thank Miss Lesley C. Gerould and Miss Suzanne Knight for performing the numerical calculations.

APPENDIX A

In spherical coordinates the solution of

$$(\nabla^2+k^2)G(r,r')=-\delta(r-r'), \quad (1)$$

where r, r' are vectors extending from the origin to the point of observation and the point source respectively, is

$$G(r,r')=\frac{e^{ik|r-r'|}}{4\pi|r-r'|}. \quad (2)$$

Physically, (2) represents spherical waves emanating from the point $r=r'$. In cylindrical coordinates (1) becomes

$$(\nabla^2+k^2)G(\rho,\rho',\phi-\phi',z-z')=\frac{-\delta(\rho-\rho')}{\rho}\delta(\phi-\phi')\delta(z-z'). \quad (3)$$

It has been shown in Sec. IV, Eq. (13) of reference 3 that (3) has the solution

$$G(\rho,\rho',\phi-\phi',z-z')=\frac{i}{8\pi}\int_{-\infty}^{\infty}\sum_{m=-\infty}^{m=\infty}\exp[im(\phi-\phi')]H_m^{(1)}\times[(k^2-\zeta^2)^{\frac{1}{2}}\rho_>]J_m[(k^2-\zeta^2)^{\frac{1}{2}}\rho_<]\exp[i\zeta(z-z')]d\zeta, \quad (4)$$

where the integration path is shown in Fig. 2. Since (2) and (4) are alternative representations, we have

$$\frac{e^{ik|r-r'|}}{4\pi|r-r'|}=\frac{i}{8\pi}\int_{-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\exp[im(\phi-\phi')]H_m^{(1)}\times[(k^2-\zeta^2)^{\frac{1}{2}}\rho_>]J_m[(k^2-\zeta^2)^{\frac{1}{2}}\rho_<]\exp[i\zeta(z-z')]d\zeta, \quad (5)$$

where $|r-r'|=(\rho^2+\rho'^2-2\rho\rho'\cos(\phi-\phi')+(z-z')^2)^{\frac{1}{2}}$. Setting $\phi'=0$ and integrating from $\phi=0$ to $\phi=2\pi$ after multiplication by

[4] E. O. Hartig, thesis, Harvard University, 1950.

$\cos\phi$, we have

$$\frac{1}{4\pi}\int_0^{2\pi}\frac{e^{ik|r-r'|}}{|r-r'|}\cos\phi d\phi=\frac{i}{4}\int_{-\infty}^{\infty}H_1^{(1)}\times[(k^2-\zeta^2)^{\frac{1}{2}}\rho_>]J_1[(k^2-\zeta^2)^{\frac{1}{2}}\rho_<]\exp[i\zeta(z-z')]d\zeta, \quad (6)$$

since

$$\int_0^{2\pi}\cos m\phi\,\cos n\phi d\phi=\pi\delta_{nm}.$$

By definition

$$\Gamma(\rho,\rho',z,z')=\frac{1}{4\pi}\int_0^{2\pi}d\phi\cos\phi\times\left\{\frac{\exp\{ik[R^2+(z-z')^2]^{\frac{1}{2}}\}}{[R^2+(z-z')^2]^{\frac{1}{2}}}-\frac{\exp\{ik[R^2+(z+z')^2]^{\frac{1}{2}}\}}{[R^2+(z+z')^2]^{\frac{1}{2}}}\right\}, \quad (7)$$

where $R^2=\rho^2+\rho'^2-2\rho\rho'\cos\phi$. Owing to Sommerfeld's representation[5]

$$\frac{\exp\{ik[R^2+(z\pm z')^2]^{\frac{1}{2}}\}}{[R^2+(z\pm z')^2]^{\frac{1}{2}}}=\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}\times\exp[-(\zeta^2-k^2)^{\frac{1}{2}}|z\pm z'|]J_0(R\zeta), \quad (8)$$

where the integration is along the real axis of the complex ζ-plane with a downward indentation at $\zeta=+k$ and $\arg(\zeta^2-k^2)^{\frac{1}{2}}=0$, $-\pi/2$ for $\zeta>k$, $\zeta<k$ respectively, we can rewrite (7) in the following form:

$$\Gamma(\rho,\rho',z,z')=\frac{1}{4\pi}\int_0^{2\pi}d\phi\cos\phi\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}J_0(R\zeta)\times\{\exp[-(\zeta^2-k^2)^{\frac{1}{2}}|z-z'|]-\exp[-(\zeta^2-k^2)^{\frac{1}{2}}|z+z'|]\}. \quad (9)$$

Then,

$$\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho,\rho',z,z')=-\frac{1}{4\pi}\int_0^{2\pi}d\phi\cos\phi\int_0^{\infty}\zeta(\zeta^2-k^2)^{\frac{1}{2}}J_0(R\zeta)\times\{\exp[-(\zeta^2-k^2)^{\frac{1}{2}}|z-z'|]+\exp[-(\zeta^2-k^2)^{\frac{1}{2}}|z+z'|]\}. \quad (10)$$

At $z=z'=0$,

$$\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho,\rho',0,0)=-\frac{1}{2\pi}\int_0^{2\pi}d\phi\cos\phi\int_0^{\infty}\zeta(\zeta^2-k^2)^{\frac{1}{2}}J_0(R\zeta). \quad (11)$$

Interchanging the order of integration and using the addition theorem,

$$J_0(R\zeta)=\sum_{n=0}^{\infty}(2-\delta_{0n})J_n(\zeta\rho)J_n(\zeta\rho')\cos n\phi,$$

we obtain

$$\int_0^{2\pi}d\phi\cos\phi J_0(R\zeta)=\sum_{n=0}^{\infty}(2-\delta_{0n})J_n(\zeta\rho)J_n(\zeta\rho')\int_0^{2\pi}\cos\phi\cos n\phi d\phi=2\pi J_1(\zeta\rho)J_1(\zeta\rho'). \quad (12)$$

Hence, by virtue of (12), (11) takes the form

$$\frac{\partial}{\partial z}\frac{\partial}{\partial z'}\Gamma(\rho,\rho',0,0)=-\int_0^{\infty}\zeta(\zeta^2-k^2)^{\frac{1}{2}}J_1(\zeta\rho)J_1(\zeta\rho')d\zeta. \quad (13)$$

Setting $z=z'=0$ in (8), we get

$$\frac{e^{ikR}}{R}=\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}J_0(R\zeta). \quad (14)$$

Multiplying through by $\cos\phi$ and integrating from $\phi=0$ to $\phi=2\pi$, (14) becomes

$$\int_0^{2\pi}\frac{e^{ikR}}{R}\cos\phi d\phi=\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}\int_0^{2\pi}J_0(R\zeta)\cos\phi d\phi=2\pi\int_0^{\infty}\frac{\zeta d\zeta}{(\zeta^2-k^2)^{\frac{1}{2}}}J_1(\zeta\rho)J_1(\zeta\rho'). \quad (15)$$

[5] Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1945).

APPENDIX B

We consider

$$f_n(\zeta)=\int_a^b\rho d\rho J_1(\zeta\rho)R_n(\rho),\quad n\geq 1. \quad (1)$$

To carry out the evaluation of $f_n(\zeta)$, we first must determine the normalization factor N_n for the eigenfunction $R_n(\rho)$:

$$R_n(\rho)=N_n[J_1(\lambda_n\rho)N_0(\lambda_n a)-J_0(\lambda_n a)N_1(\lambda_n\rho)], \quad (2)$$

where λ_n are roots of

$$N_0(\lambda_n a)J_0(\lambda_n b)=J_0(\lambda_n a)N_0(\lambda_n b). \quad (3)$$

N_n are chosen such that

$$\int_a^b R_n(\rho)R_m(\rho)\rho d\rho=\delta_{nm}. \quad (4)$$

Let

$$Z_p=J_p(\lambda_n\rho)N_0(\lambda_n a)-J_0(\lambda_n a)N_p(\lambda_n\rho). \quad (5)$$

Then (4) becomes

$$N_n^2\int_a^b[Z_1(\lambda_n\rho)]^2\rho d\rho=1=N_n^2\left\{\frac{\rho^2}{2}(Z_1^2(\lambda_n\rho)-Z_0(\lambda_n\rho)Z_2(\lambda_n\rho))\right\}_a^b. \quad (6)$$

With the aid of the identity $J_1(x)N_0(x)-J_0(x)N_1(x)=2/\pi x$ and (3), it follows from (6) that

$$\frac{1}{N_n}=\frac{\sqrt{2}}{\pi\lambda_n}\left[\frac{J_0^2(\lambda_n a)}{J_0^2(\lambda_n b)}-1\right]^{\frac{1}{2}}. \quad (7)$$

Substituting (2) into (1) and noting that

$$\int_a^b\rho d\rho J_1(\zeta\rho)J_1(\lambda_n\rho)=\frac{\zeta\rho J_1(\lambda_n\rho)J_0(\zeta\rho)-\lambda_n\rho J_0(\lambda_n\rho)J_1(\zeta\rho)}{\lambda_n^2-\zeta^2}\Bigg|_{\rho=a}^{\rho=b}, \quad (8)$$

and

$$\int_a^b\rho d\rho J_1(\zeta\rho)N_1(\lambda_n\rho)=\frac{\zeta\rho N_1(\lambda_n\rho)J_0(\zeta\rho)-\lambda_n\rho N_0(\lambda_n\rho)J_1(\zeta\rho)}{\lambda_n^2-\zeta^2}\Bigg|_{\rho=a}^{\rho=b}, \quad (9)$$

we finally get:

$$f_n(\zeta)=\int_a^b\rho d\rho J_1(\zeta\rho)R_n(\rho)=\frac{2}{\pi}\frac{N_n}{\lambda_n}\frac{1}{J_0(\lambda_n b)}\frac{\zeta}{\lambda_n^2-\zeta^2}[J_0(\zeta b)J_0(\lambda_n a)-J_0(\zeta a)J_0(\lambda_n b)], \quad (10)$$

when $n=0$,

$$f_0(\zeta)=\int_a^b\rho d\rho J_1(\zeta\rho)R_0(\rho)=\int_a^b d\rho J_1(\zeta\rho)=\frac{J_0(\zeta a)-J_0(\zeta b)}{\zeta}. \quad (11)$$

APPENDIX C

Sonine's formula (p. 417, Eq. 5, reference 5)

$$\int_0^{\infty}\frac{J_\nu[a(\beta^2+z^2)^{\frac{1}{2}}]}{(\beta^2+z^2)^{\nu/2}}\beta^{2\mu+1}d\beta=\frac{2^\mu\Gamma(\mu+1)}{a^{\mu+1}z^{\nu-\mu-1}}J_{\nu-\mu-1}(az), \quad (1)$$

is valid provided $a\geq 0$, $Re(\frac{1}{2}\nu-\frac{1}{4})$, $Re(\mu)>-1$, where Re denotes the real part.

If we let $\xi=(\beta^2+z^2)^{\frac{1}{2}}$ it follows from (1) that

$$\int_z^{\infty}\frac{J_\nu(a\xi)}{\xi^{\nu-1}}(\xi^2-z^2)^\mu d\xi=\frac{2^\mu\Gamma(\mu+1)}{a^{\mu+1}z^{\nu-\mu-1}}J_{\nu-\mu-1}(az). \quad (2)$$

We now consider the integral

$$F(a,b,z)\equiv\int_z^{\infty}\frac{J_0(a\xi)J_0(b\xi)}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi=\frac{1}{\pi}\int_0^{\pi}d\theta\int_z^{\infty}\frac{J_0[\xi(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}]}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi, \quad (3)$$

where use is made of Neumann's addition theorem and the

orthogonality of the trigonometric functions in passing from the first integral definition to the second.

Let

$$G(a, b, z)=\int_z^\infty \frac{J_0[\xi(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}]}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi. \tag{4}$$

Then

$$\frac{\partial}{\partial a}G(a, b, z)=-\frac{a-b\cos\theta}{(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}} \times\int_z^\infty \frac{J_1[\xi(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}]}{(\xi^2-z^2)^{\frac{1}{2}}}d\xi, \tag{5}$$

with the aid of (2) [setting $\nu=1$, $\mu=-\frac{1}{2}$, and recalling that $J_{\frac{1}{2}}(x)=(2/\pi x)^{\frac{1}{2}}\sin x$] (5) becomes

$$\frac{\partial}{\partial a}G(a, b, z)=-\frac{a-b\cos\theta}{a^2+b^2-2ab\cos\theta}\cdot\frac{1}{z}\cdot\sin[z(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}]. \tag{6}$$

Integrating (6), we get

$$G(a, b, z)-G(0, b, z) =-\frac{1}{z}\int_0^a \frac{x-b\cos\theta}{x^2+b^2-2bx\cos\theta}\sin[z(x^2+b^2-2xb\cos\theta)^{\frac{1}{2}}]dx. \tag{7}$$

Making a change of variable, $v=(x^2+b^2-2xb\cos\theta)^{\frac{1}{2}}$, (7) takes the form

$$G(a, b, z)-G(0, b, z)=\frac{1}{z}Si(bz)-\frac{1}{z}Si[z(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}], \tag{8}$$

where

$$Si(x)=\int_0^x \frac{\sin t}{t}dt.$$

To evaluate $G(0, b, z)$ we note that by (4)

$$G(0, b, z)=\int_z^\infty \frac{J_0(\xi b)}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi. \tag{9}$$

then

$$\frac{\partial}{\partial b}G(0, b, z)=-\int_z^\infty \frac{J_1(\xi b)}{(\xi^2-z^2)^{\frac{1}{2}}}d\xi, \tag{10}$$

or with the aid of (2) (setting $\nu=1$, $\mu=-\frac{1}{2}$)

$$\frac{\partial}{\partial b}G(0, b, z)=-\frac{1}{bz}\sin bz. \tag{11}$$

Integrating (11) with respect to b, we get

$$G(0, b, z)-G(0, 0, z)=-\frac{1}{z}\int_0^b \frac{\sin bz}{b}db=-\frac{1}{z}Si(bz). \tag{12}$$

But since

$$G(0, 0, z)=\int_z^\infty \frac{d\xi}{\xi(\xi^2-z^2)^{\frac{1}{2}}}=\frac{\pi}{2z}, \tag{13}$$

we obtain

$$G(0, b, z)=\frac{\pi}{2z}-\frac{1}{z}Si(bz), \tag{14}$$

which when substituted into (8) gives

$$G(a, b, z)=\frac{\pi}{2z}-\frac{1}{z}Si[z(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}]. \tag{15}$$

It follows from (3), (4), and (15) that

$$\int_z^\infty \frac{J_0(a\xi)J_0(b\xi)}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi=\frac{\pi}{2z}-\frac{1}{\pi z}\int_0^\pi Si[z(a^2+b^2-2ab\cos\theta)^{\frac{1}{2}}]d\theta, \tag{16}$$

and when $b=a$

$$\int_z^\infty \frac{J_0^2(a\xi)}{\xi(\xi^2-z^2)^{\frac{1}{2}}}d\xi=\frac{\pi}{2z}-\frac{1}{\pi z}\int_0^\pi Si\left(2za\sin\frac{\theta}{2}\right)d\theta. \tag{17}$$

Digital Information Storage in Three Dimensions Using Magnetic Cores*

JAY W. FORRESTER
Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 5, 1950)

Present digital storage devices use two space coordinates or time and one space coordinate for selection switching, resulting in bulky construction or long access time. Three-dimensional arrays with efficient high speed selection appear possible after continued development of rectangular-hysteresis magnetic materials. An operating mode is suggested which depends on ability of the magnetic material to discriminate between two values of magnetizing force which differ by a 2:1 ratio. Only one magnetic core per binary digit is required. Tests show that most existing metallic magnetic materials switch in 20 to 10,000 microseconds and are too slow. Nonmetallic magnetic materials can now approach the required magnetic behavior; they switch in less than a microsecond.

INTRODUCTION

ALL digital storage devices currently used in electronic computers have serious shortcomings, and we should expect major improvements in the future. Such improvements will probably come through new combinations of storage elements and switching systems. This article discusses one such possibility.

The storage of digital information is more a problem of selection and switching than it is a problem of information retention. Many simple physical devices are available to store information, but most of them lack a suitable high speed selecting system. The best storage and selecting systems in present use can be divided into two types: those in which time is used as one of the selecting dimensions, and those which make a selection on the basis of two space coordinates. The acoustic delay line and the magnetic drum use time as one of the selecting dimensions; consequently, the storage systems have a relatively low access speed. The various electrostatic storage tubes select information on the basis of two space coordinates which can be rapidly controlled. Electrostatic tubes are, however, rather expensive, possess a relatively short life, and, compared with an ideal system, are bulky and awkward.

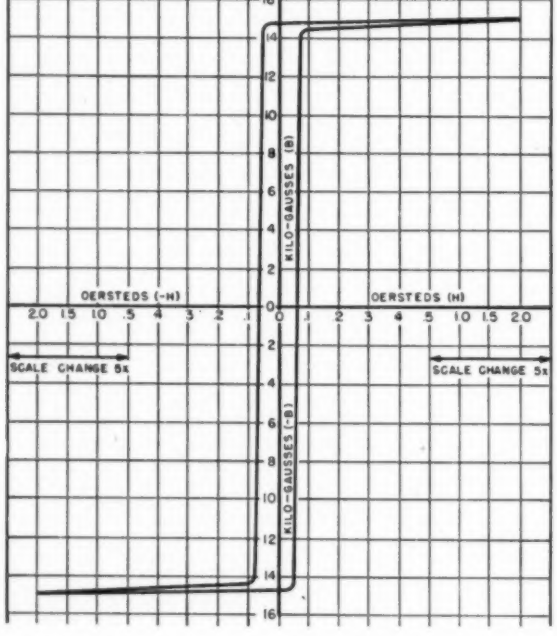


FIG. 1. Hysteresis loop for Deltamax made by Allegheny Ludlum Steel Corporation and obtained from Arnold Engineering Company, Chicago.

In an ideal storage system it should be possible to arrange elementary storage cells in a compact three-dimensional array.[1] Storage elements inside the volume should be selected by suitably controlling three coordinates along the edges of a solid array.

This article discusses the possibility of a three-dimensional array of storage elements using magnetic cores having rectangular hysteresis loops. Magnetic materials with rectangular hysteresis loops have been used for digital storage at the Harvard Computation Laboratory and elsewhere.[2] In existing storage systems, the magnetic cores have been used in either of two ways: (1) for the storage of isolated digits, where selection was not a basic problem; or (2) arranged in the form of delay lines in which information is stepped from one core to another, and time is again used as one of the selecting dimensions.

The method herein proposed has thus far received only a preliminary evaluation on the basis of studies on individual magnetic cores.

RECTANGULAR HYSTERESIS LOOPS

Magnetic materials are available having a nearly rectangular hysteresis loop as shown in Fig. 1. Figure 2 shows an idealized form of the curve which will be used in the following discussion. For magnetizing forces greater than H_c, the material reaches a sharply defined knee in the magnetization curve, above which very little change in magnetic flux occurs. After a high positive H is applied and then removed, the material will reside in the condition A of Fig. 2. Likewise, after

* This research is a part of Project Whirlwind sponsored by the ONR under Contract N5ori-60.

[1] Jay W. Forrester, "Data storage in three dimensions," 13 pages, Project Whirlwind Report No. M-70, the M.I.T. Servomechanisms Laboratory (April 29, 1947).

[2] An Wang and Way Dong Woo, J. Appl. Phys. **21**, 49–54 (1950).

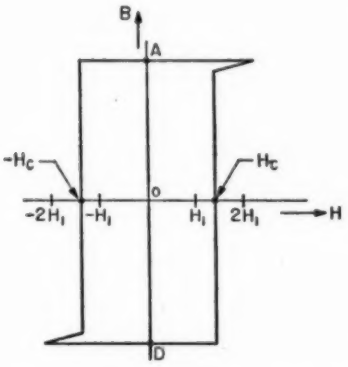


FIG. 2. Idealized hysteresis loop showing 2:1 ratio of magnetizing forces used in switching.

a high negative H, the material will reside at condition D. The two conditions A and D, corresponding to the storage of binary digits 0 and 1, will be retained by the core until a magnetizing force exceeding H_c is applied. A detecting coil on the magnetic core can be used to determine its condition. When a strong positive H is applied to an exciting coil, the detecting coil will show a very small voltage if the core is in condition A and a large voltage if the core is in condition D. The large signal corresponding to condition D arises from the flux linkages through the detecting coil when the magnetic condition of the core is reversed from D to A.

Existing applications of rectangular-hysteresis-loop materials to digital storage have depended on the behavior of the core at zero magnetizing force and at high positive and negative values of H. The storage system described in this report is dependent on the response of the core to positive and negative values of H_1 and positive and negative values of $2H_1$ as indicated in Fig. 2. The material must be sufficiently rectangular that when the core is at condition D a magnetizing force H_1 will cause little change in flux density; yet twice this value, or $2H_1$, will cause almost complete reversal of the magnetic flux to condition A. The switching system is based on the ability of the core to discriminate between H_1 and $2H_1$. This ability has been demonstrated in the laboratory.

TWO-DIMENSIONAL STORAGE ARRAY

A three-dimensional storage array is best understood by first examining the two-dimensional array of Fig. 3. The magnetic cores are presumed to have hysteresis characteristics similar to Fig. 2. The x and y circuits are used for selecting cores both for the reading and writing processes and can be individually excited with currents corresponding to the positive or negative H_1 values of Fig. 2.

Consider a writing operation in which positive current is circulated through circuits x_1 and y_1. Core C_{11} will be excited by two coils each carrying a current corresponding to H_1. The resulting $2H_1$ is sufficient to switch the magnetic material to condition A; and, by definition, the binary digit one has been writen in core C_{11}. During this operation cores C_{12} and C_{21} have each been excited by a single current corresponding to H_1 which is insufficient to cause a change in their magnetic state. Core C_{22} has received no excitation. Correspondingly, we can circulate a negative current through circuits x_1 and y_1 and change core C_{11} back to state D of Fig. 2, again without affecting the other three cores.

An output circuit for reading can be arranged in many ways, one of which is shown in Fig. 3. Sensing coils detect the output signal, and the reading operation uses x and y currents just as for the writing-positive operation. During reading, if core C_{11} is in the D state, it will be triggered by the positive currents through x_1 and y_1. The flux linkages through the detecting coil will give an output signal in the output mixer circuit. On the other hand, if core C_{11} is already in the A condition, very little signal will be generated in the output circuit when the exciting currents are applied. The output sensing system must thus discriminate between the magnitudes of these two signals for determining whether the core is storing a binary 0 or a binary 1. Obtaining an output signal erases condition D of Fig. 2. Following a reading operation it is, therefore, necessary to re-establish condition D in core C_{11} if further reading is desired.

The two-dimensional array of Fig. 3 makes it possible to select storage cores by means of a number of leads:

$$\text{Selecting lines} = 2\ (\text{no. of cores})^{\frac{1}{2}} = 2K^{\frac{1}{2}},$$

where K equals the total number of storage cells.

To be practical, the system must have a physical design such that many thousands of magnetic cores can be arranged simply and inexpensively. This could not be accomplished if the cores were wound with three separate coils as implied in Fig. 3. With suitable transforming of power levels the cores can be operated with single-turn coils. A single turn is by definition a wire running once through the center of the magnetic core as shown in Fig. 4a. The straight conductors x and y in the plane of the paper correspond to the x and y circuits of Fig. 3. The arrangement of Fig. 3 could then be constructed from a set of straight crossed connectors

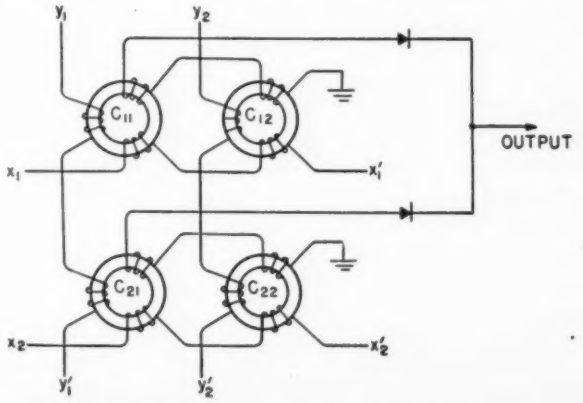


FIG. 3. Two-dimensional storage and switching array.

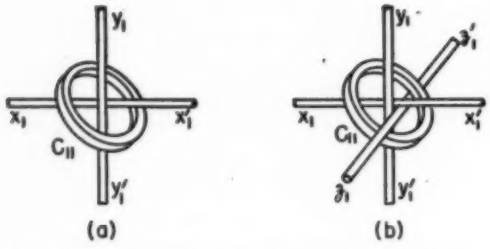


FIG. 4. Single-turn windings in perpendicular planes.

with magnetic cores placed around their intersections at a 45° angle.

As shown in Fig. 4b, it is possible to insert three straight mutually perpendicular conductors through the center of a magnetic core whose plane is at approximately 45° to each of the three axes.

THREE-DIMENSIONAL STORAGE ARRAY

We are now ready to consider a three-dimensional array of magnetic cores. In Fig. 5 each elementary cube of the three-dimensional array is presumed to contain a magnetic core surrounding three perpendicular conductors as in Fig. 4b. Only a few of the straight electrical conductors are shown. All vertical conductors in the plane z_1 are connected in parallel. Likewise, all horizontal conductors in plane z_1 are connected in parallel through suitable mixing circuits to the single output line. Horizontal conductors perpendicular to the xy plane and parallel to the z axis are individually selected for the purposes of this present explanation. The vertical circuits connecting z_1 and z_1' might better be connected in series rather than parallel, depending on the nature of the driving circuit.

Suppose now that a current is circulated between z_1 and z_1' of such magnitude that each magnetic core in the z_1 plane receives a magnetizing force equal to H_1 of Fig. 2. This is insufficient to create any change in the state of magnetization of the cores. We can at the same time select one of the wires extending through the xy plane, for example, the conductor x_1y_1. By circulating a current from x_1y_1 to $x_1'y_1'$, it is now possible as before in Fig. 3 to obtain a magnetizing force equal to $2H_1$ at the magnetic core C_{111} and at no other core. Positive currents will create one state of magnetization, while negative currents will create the other.

In a manner similar to that described for Fig. 3, the horizontal output circuits can detect magnetic switching of core C_{111} and use this as an output signal.

Through selection of the external circuits as illustrated it is possible to choose any core within the solid array and to determine its existing state of magnetization or to establish it in either of its two desired states.

SELECTION SYSTEM

Switching as illustrated in Fig. 5 requires the selection of one of the conductors in the xy plane as well as selection of one of the z planes. Assuming n values each along the x and y axes and m values along the z axis, the number of circuits to be selected is:

Selecting lines $= n^2+m$.
Total digits stored $= n^2m$.
For a specified number of digits, K,

$$n^2m = K$$
$$n^2 = K/m.$$

Selecting lines $= K/m+m$, and the minimum number of selecting lines occurs when $m = K^{\frac{1}{2}}$.
Minimum selecting lines $= 2K^{\frac{1}{2}}$ as before in the two-dimensional example.

Although the storage elements have now been arranged in three-dimensional array, there has been no reduction in selecting lines; and it will be necessary to have a more efficient selecting mechanism so that fewer control lines need be handled.

Figure 6 shows a method whereby the selection of a z plane, an x plane, and a y plane is sufficient for locating a core at their intersection. Only the storage elements in the z_1 plane are shown. A second horizontal set of wires parallel to the first set and perpendicular to the xy plane is used as shown. Four conductors now go through each magnetic core. Three of the conductors correspond to the x, y, and z planes, and the fourth is the output signal circuit. The connection in the z plane and the connections for the signal circuits are the same as in Fig. 5. The far ends of the x- and y-plane circuits are not shown. Assume now that core C_{111} is to be selected. Circuit x_1 will energize all cores in the top plane of the cube at a value of magnetizing force equal to H_1 in Fig. 2. Assume now that a *negative* current is established in all y planes *except* y_1. In other words, y_2, y_3, and y_4 are each energized with $-H_1$ magnetizing force. Assume now that plane z_1 is energized to produce a positive magnetizing force H_1. Core C_{111} is now subjected to a magnetizing force $2H_1$, while all other cores in the cube are being subjected to either zero excitation

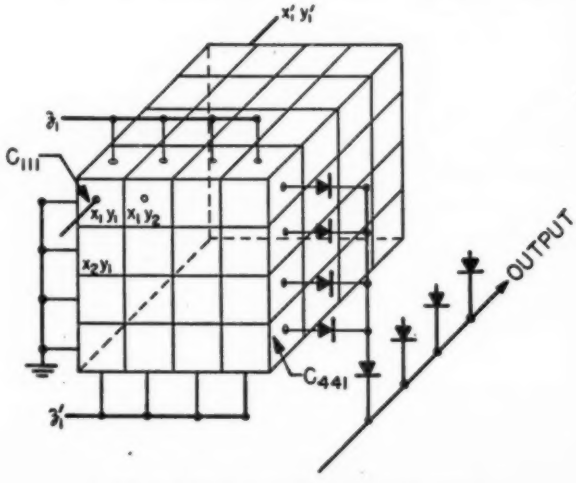


FIG. 5. Three-dimensional storage array.

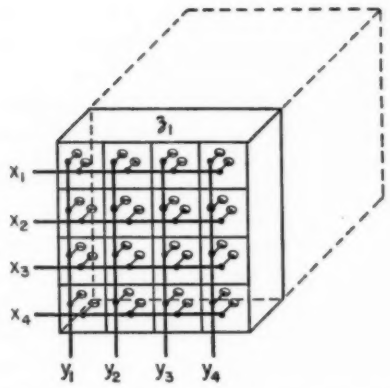


FIG. 6. Three-dimensional storage array with coordinate plane selection.

or to an excitation of plus or minus H_1. Core C_{111} has the combined excitations of planes x_1 and z_1. The cores below C_{111} are subjected to only the z-plane excitation. Cores to the right of C_{111} have x- and y-plane currents cancelling each other, the net result being the z-plane current. The remaining cores in the z_1 plane are subjected to zero magnetizing force, since the y- and z-plane currents cancel each other. Likewise, in the remainder of the cube the cores are subjected either to the individual x or y excitations or to zero excitation. The number of selecting lines required to select the intersection of 3 planes in this manner is:

$$\text{Selecting lines} = 3K^{\frac{1}{3}}.$$

For the amounts of storage currently contemplated in high speed computing machines, the selection of this number of control leads is well within the present standard practice for matrix switching devices and presents no particular problem.

PRACTICABILITY

The principles of storage and switching having been established, there remains a discussion of the possibility of physically executing the system. Economically, the system is probably well within reason, since at a cost of about one dollar or less per magnetic core it is competitive with the initial cost of electrostatic storage tubes. There remains the question of whether adequate speeds can be reached and whether or not with a practical signal-pick-up circuit the output of the selected coil can be distinguished from switching noise which will occur in other cores leading into the same detector system. The proper choices of coordinate directions, parallel and series connections, orientation of cores on the crossed conductors, and order of impressing currents will reduce access time and increase the output signal-to-noise ratio.

It should be noted that the z planes of Fig. 5 or 6 can correspond to the separate digit columns of a parallel-type computer, and simultaneous operation of the separate digits is possible.

PRELIMINARY EXPERIMENTAL RESULTS

In this section no attempt is made to give complete test results. Recently a thesis research study has given additional valuable and encouraging information on magnetic core behavior.[3]

As seen in Fig. 2, the core must operate under conditions such that at a magnetizing force H_1 essentially no switching of the core takes place, while at a magnetizing force $2H_1$, the core must switch rapidly from one state of magnetization to the reverse.

Figure 7 shows a study of reversal time as a function of driving-pulse ampere-turns for a Deltamax core. In this core the direct-current coercive force is equivalent to about 0.3 ampere-turn. No switching of the core takes place below that driving current. Speed increases with driving force as shown on log-log paper. Core reversal can be accomplished in about two microseconds with a driving force of some 40 ampere-turns, which is more than 100 times the dc coercive force. At a driving current of twice the dc coercive force, the switching time of the core is approximately 10,000 microseconds. This speed is far removed from the few microseconds that an ideal system should exhibit. The curve for continuous square-wave excitation departs at higher frequencies from the curve for occasional isolated switch-

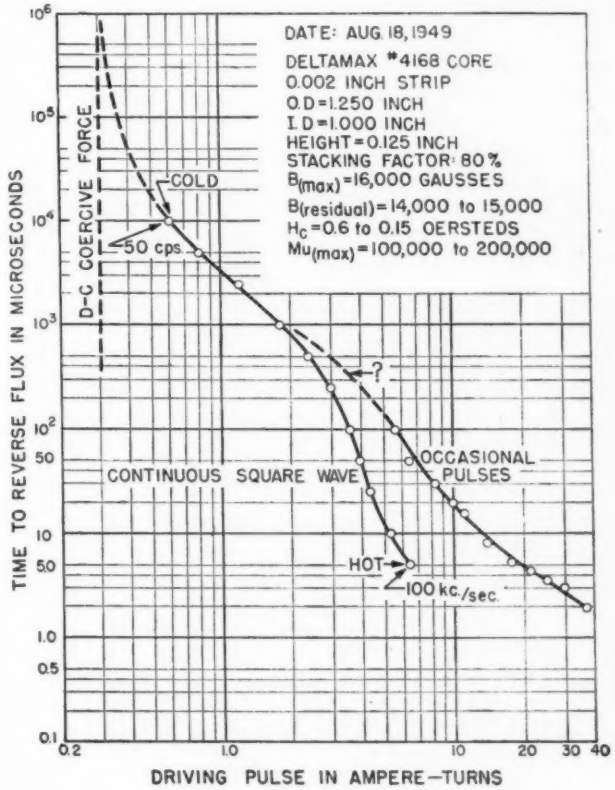


FIG. 7. Reversing speed with increasing amplitudes of rectangular magnetizing force pulses.

[3] William N. Papian, "A Coincident Current Magnetic Memory Unit," Master of Science thesis, M.I.T. Electrical Engineering Department, 1950.

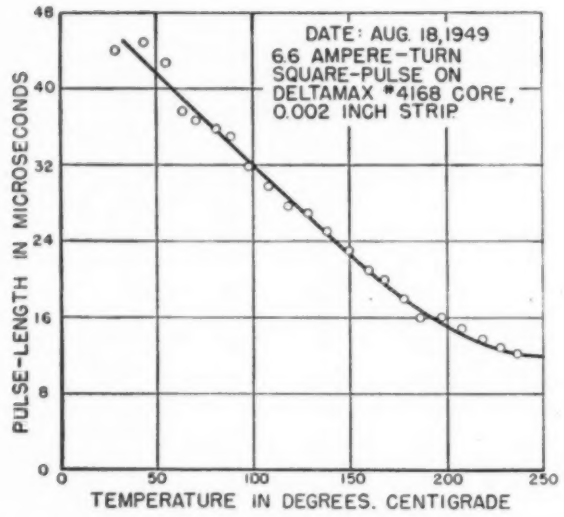


FIG. 8. Reduction in switching time as temperature is increased.

ing pulses, probably because of the temperature rise created by the iron losses in the continuous switching cycle.

Figure 8 is an interesting curve showing the dependence of switching time on core temperature. This is the same size core as in Fig. 7 driven with a 6.6-ampere-turn square pulse. Switching time declines from 44 microseconds at about 40 degrees Centigrade to 12 microseconds at 250 degrees Centigrade. There is some evidence that irreversible damage may occur to the cores at the highest temperatures.

The Delatmax material is unsuitable, at least on the basis of these tests, for use in a high speed switching system in which exciting current must not exceed twice the coercive force. Other cores with higher values of H_c have been given preliminary tests which indicate a switching time of 30 or 40 microseconds for values of magnetizing force equal to $2H_1$. High switching speed not only gives faster operation but also higher output-signal levels for a given core size. The core of Fig. 7 gives a 6-volt output in a one-turn loop when driven to reversal in two microseconds.

Cores wound from thinner material speed up the switching action through a decrease in eddy-current shielding within the material. Switching time is not, however, proportional to the square of thickness. The speed ratio improvement appears to decrease at the higher speeds and this, as well as the dependence on temperature, seems to indicate that a phenomenon other than eddy current shielding is limiting the high speed operation.

To obtain higher speed operation, future work will investigate the use of thinner core materials, the use of core materials having a higher coercive force, materials with a higher resistance, and especially the nonmetallic magnetic materials.

Materials with a higher coercive force are promising. The switching speed is probably a function of the number of ampere-turns applied in excess of the coercive force. Higher coercive forces permit higher values of H_1 and correspondingly higher values of $2H_1$. With higher values of H_c, there can then be a greater numerical difference between H_c and $2H_1$ than for materials with low coercive force. This is verified in experimental tests.

The nonmetallic magnetic ferrites are especially interesting because of their high volume resistance, their potentially high operating speed, and their low cost of fabrication. Little seems to be known about obtaining rectangular hysteresis loops in these materials. Even some commercially available nonmetallic materials come close to meeting the requirements of the selecting system and at the same time can be switched in a fraction of a microsecond.

Also of great future promise in three-dimensional storage arrays are the ferroelectric materials which make possible condensers with rectangular hysteresis response and nonlinear characteristics comparable to magnetic materials. With the aid of simple resistance networks, coordinate switching can follow the general scheme outlined above. Ferroelectric materials may permit higher speeds and more convenient driving impedances than ferromagnetic storage.

ERRATUM: Erratum to Mechanical Properties of Glass

[J. App. Phys. 21, 1195 (1950)]

F. W. PRESTON
Preston Laboratories, Butler, Pennsylvania

Through an error in the editorial office, a second paragraph was added to an erratum for the original paper by F. W. Preston which appeared in the October, 1942, issue of the Journal of Applied Physics. This second paragraph is obviously in error and the velocity of sound in glass is about 3 miles per second as was stated on page 624 of the original paper.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Parametric Excitation

N. MINORSKY
Stanford University Office of Naval Research, Stanford University, California
(Received June 6, 1950)

This paper outlines a theory of excitation of oscillations under the rather slow variations of a parameter on which a system depends. The basic equation is a Mathieu equation. The basic method consists in reducing the system to polar coordinates and applying perturbations.

1. INTRODUCTION

IT has been known for a long time that if a parameter of an oscillatory system (electrical or mechanical) is varied periodically between certain limits, the system becomes "excited," that is, starts oscillating with frequency equal to one-half of that with which the parameter varies. The term "parametric excitation" (p.e. for short) is frequently used to designate this phenomenon.

Perhaps the best known example of p.e. is the operation of a swing, that is, the "excitation" of a swing due to a properly timed bending of the knees followed by a subsequent straightening out on the part of the person on the swing. Owing to this periodic raising or lowering of the center of gravity of the body, the motion of the swing starts from a small angle and is subsequently maintained at a considerable angle of oscillation.

Melde[1] and Lord Rayleigh[2] were first to observe and analyze this phenomenon. Rayleigh's experimental arrangement consisted of a stretched wire, one end of which was fastened to the prong of a tuning fork. When the fork is set in oscillation, thus causing a variation of tension in the string, lateral oscillations of the latter are excited with a frequency equal to one-half of that of the tuning fork.

In later years L. Mandelstam and N. Papalexi[3] constructed an interesting "parametric generator" consisting of an oscillatory circuit with a variable parameter (either the capacity C or inductance L). It is observed that when the parameter is made to vary around its average value with frequency f, the circuit begins to oscillate with frequency $f/2$ in the *absence of any external source of electric energy*. The experiment is particularly well defined when the frequency $f/2$ coincides with the natural frequency of the oscillating circuit. In their early experiments these authors operated with a nearly linear circuit, in which case it was observed that the voltage generated in the apparatus rapidly reaches such a high value that the machine is punctured. In order to obviate this, a nonlinear resistor was inserted in series with the circuit; and it was found that the voltage (and hence the energy generated in the device) approaches a definite stationary value.

The study of this phenomenon is obviously amenable to the differential equation (d.e. for short) with periodic coefficients, a subject on which there exists a voluminous literature for the last seventy years or so. It is well known that a typical equation of this kind, the so-called Mathieu equation, has stable and unstable zones in its "parameter space." The p.e., by its nature, is to be expected in the latter. The experiment shows, however, that the phenomenon always adjusts itself so as to be in these unstable zones. There is another complication, namely, the lack of any theory of these d.e. in the nonlinear range. On the other hand, most frequently these phenomena occur precisely when the system is nonlinear. In what follows we propose to outline a theory of this phenomenon on the basis of the perturbation method introducing the polar coordinates $[\rho(t), \theta(t)]$ where $\rho(t)$ will be connected with the total energy and $\theta(t)$ will be the phase angle of the motion. It will be shown that the existence of this effect is closely related to the question of *stability of the phase* around such a value, for which a steady increase of energy is possible. Moreover, a nonlinear extension of this method is also relatively simple.

2. LINEAR CASE; STABILITY OF THE PHASE

Since the reduction of the various practical cases of p.e. to the Mathieu equation is well known,[4] we shall start from the d.e. of the form

$$\ddot{x}+(1+\gamma\cos\omega t)x=0, \qquad (2.1)^*$$

where x is the variable of the problem (e.g., coordinate, voltage, etc.), γ is the so-called index of modulation, and ω is the frequency with which the parameter is varied.

Equation (2.1) is equivalent to the system of

[1] F. E. Melde, Pogg. Annalen **108**, 508 (1859).
[2] Lord Rayleigh, "On maintained vibrations," Phil. Mag. **15**, 229–235, Fifth Series (April, 1883).
[3] L. Mandelstam, N. Papalexi, A. Andronov, A. Witt, and S. Chaikin, "Exposé des recherches recentes sur les oscillations nonlinéaires," J. Tech. Phys. (U.S.S.R.) **2**, 81–134 (1934) (a bibliography on parametric excitation is appended to this paper).

[4] N. W. McLachlan, *Theory and Applications of Mathieu Functions* (Oxford University Press, 1947), Sec. 15.50–15.54.

* It is preferable to deal throughout with dimensionless quantities. If one designates time by t', angular velocity by ω', the usual form of the Mathieu equation encountered in applications is $d^2x/dt'^2+(\alpha'+\beta'\cos\omega' t')x=0$. Here α' and β' are obviously of dimension T^{-2}. Change of the independent variable from t' to $t=(\alpha')^{\frac{1}{2}}\cdot t'$ reduces this equation to (2.1), where $\gamma=\beta'/\alpha'$, $\omega=\omega'(\alpha')^{\frac{1}{2}}$ and $t=(\alpha')^{\frac{1}{2}}\cdot t'$ are clearly dimensionless. In Eq. (7.1) we use as independent variable t' (time), but beginning with (7.2) we deal again with the dimensionless time $t=\omega_0 t'$.

equations

$$\dot{x}=y$$
$$\dot{y}=-(1+\gamma\cos\omega t)x.$$

Changing to polar coordinates, that is, setting

$$x=r\cos\theta \qquad \rho=r^2=x^2+y^2$$
$$y=r\sin\theta \qquad \theta=\arctan(y/x),$$

the above system becomes

$$d\rho/dt=-\gamma\rho\cos\omega t\sin 2\theta, \tag{2.2}$$

$$d\theta/dt=-1-\gamma\cos\omega t\cos^2\theta. \tag{2.3}$$

The variable $\rho=x^2+\dot{x}^2=x^2+y^2$ may be regarded as *the square of the* solution $r(\theta)$ in the phase plane (r, θ). For a conservative system one can identify the curve $\rho(\theta)$ with the total energy of the system which is then constant. However, this does not apply here, since the system is not conservative. One may retain a certain connection with energy if one agrees to consider the motion of the representative point on the $\rho(t)$ curve as evolution of energy of the system in time. Thus, for example, if $\rho(t)\to\infty$ when $t\to\infty$, this will mean that the energy grows beyond any bound. If, however, $\rho(t)$ describes a certain closed curve $\rho(\theta)$, this will mean that a certain stationary (not necessarily constant in the instantaneous value) energy content of the system is ultimately reached. The case $\omega=2$ is of primary interest, since precisely in this case p.e. has been observed experimentally. This gives

$$d\rho/dt=-\gamma\rho\cos 2t\sin 2\theta \tag{2.4}$$

$$d\theta/dt=-1-\gamma\cos 2t\cos^2\theta. \tag{2.5}$$

Introducing the variable $P(t)=\log\rho(t)$ and choosing any arbitrary instant of motion as $t=0$, one can set $P(0)=P_0$; $\theta(0)=\varphi_0$. The solutions of Eqs. (2.4) and (2.5) are determined uniquely and, since they depend analytically on the parameter γ, one can represent them by series development:

$$P(t)=\sum_{\nu=0}^{\infty}\gamma^\nu P_\nu(t); \quad \theta(t)=\sum_{\nu=0}^{\infty}\gamma^\nu\theta_\nu(t). \tag{2.6}$$

If one introduces these series into Eqs. (2.4) and (2.5) and equates the terms with equal powers of γ, one obtains a series of d.e.,

$$\dot{P}_0=0, \quad \dot{P}_1=-\cos 2t\sin 2\theta_0;$$
$$\dot{P}_2=-2\cos 2t\cos 2\theta_0\cdot\theta_1; \tag{2.7}$$

$$\dot{\theta}_0=-1; \quad \dot{\theta}_1=-\tfrac{1}{2}\cos 2t.\ (1+\cos 2\theta_0);$$
$$\dot{\theta}_2=\cos 2t.\ \sin 2\theta_0.\ \theta_1. \tag{2.8}$$

From these one can compute recurrently all terms of the expansion (2.6) with the initial conditions implied by the set-up, *viz.*:

$$P_\nu(0)=0; \quad \theta_\nu(0)=0, \quad \nu=1, 2, \cdots. \tag{2.9}$$

In this manner one obtains

$$P_0(t)=P_0; \quad P_1(t)=-(\tfrac{1}{2}\sin 2\varphi_0)t$$
$$+\text{periodic terms}\cdots \tag{2.10}$$

$$\theta_0(t)=\varphi_0-t; \quad \theta_1(t)=-(\tfrac{1}{4}\cos 2\varphi_0)t$$
$$+\text{periodic terms}, \cdots, \tag{2.11}$$

where the periodic terms resume the same values if t is increased by 2π.

Thus, starting from a point (P_0, φ_0) at $t=0$ in the phase plane, one reaches at $t=2\pi$ the point $[P(2\pi), \theta(2\pi)]$ given by the expressions

$$P(2\pi)=P_0-\pi\gamma\sin 2\varphi_0+0(\gamma^2) \tag{2.12}$$

$$\theta(2\pi)=\varphi_0-2\pi-\tfrac{1}{2}\pi\gamma\cos 2\varphi_0+0(\gamma^2), \tag{2.13}$$

where the terms $(0(\gamma^2)$ are analytic functions of φ_0 and γ and are independent of P_0. These terms result from the secular terms of the higher orders.

It is to be noted that the values P_0 and φ_0 enter into the solutions (2.10) and (2.11) in a different fashion. It is clear from Eqs. (2.4) and (2.5) that with each solution $P(t)$ and $\theta(t)$, the functions $P(t)+\text{const}$, $\theta(t)$ provide another solution. Thus, P_0 enters the solution in a simple additive way, while the dependence of the solution on φ_0 is more complicated.

It is seen that in the time interval 2π the point (P_0, φ_0) undergoes transformations of the form (2.12) and (2.13). Consider first the changes of the phase only modulo 2π; in each period 2π the phase will vary by an amount

$$\Delta\varphi_0=-\tfrac{1}{2}\pi\gamma\cos 2\varphi_0+0(\gamma^2). \tag{2.14}$$

There are two distinct values of φ_0 leading to $\Delta\varphi_0=0$; namely,

$$\varphi_0'=(\pi/4)+0(\gamma) \quad \text{and} \quad \varphi_0''=(3\pi/4)+0(\gamma), \tag{2.15}$$

as easily observed from (2.14) and the requirement $\Delta\varphi_0=0$.

For $\gamma\ll 2\pi$ in the neighborhood of φ_0' one has

$$\Delta\varphi_0>0 \text{ for } \varphi_0>\varphi_0'; \quad \Delta\varphi_0<0 \text{ for} \varphi_0<\varphi_0'. \tag{2.16}$$

For φ in the neighborhood of φ_0'' one has

$$\Delta\varphi_0<0 \text{ for } \varphi_0>\varphi_0''; \quad \Delta\varphi_0>0 \text{ for } \varphi_0<\varphi_0''. \tag{2.17}$$

In other words, the point φ_0' is a point of repulsion for neighboring phases under the transformation (2.14), while the point φ_0'' is a point of attraction. This means that φ_0'' is a stable phase which any initial phase will approach eventually.

While the initial phase φ_0 approaches its stable value φ_0'', changes will occur also in the energy term P. Once the stable phase has been reached, the energy term increases in each interval 2π by a fixed positive amount,

$$\Delta P_0=+\pi\gamma+0(\gamma^2), \tag{2.18}$$

so that the energy ultimately grows beyond any bound as was actually observed by Mandelstam and Papalexi in their experiments with a linear parametric generator.

3. FREQUENCIES OTHER THAN $\omega=2$

All experiments evidence so far available relates to the case when $\omega=2$ in the d.e. (2.2) and (2.3). In com-

parison, the question as to whether p.e. may occur for other values of ω has been neglected. It can be shown, however, that, at least as far as the terms of the first order are concerned, the p.e. does not appear for any frequency other than $\omega=2$.

Proceeding as previously one finds

$$\theta_1=-(1/2\omega)\sin\omega t-[1/4(\omega+2)]\cos2\varphi_0\sin(\omega+2)t -[1/4(\omega-2)]\cos2\varphi_0\sin(\omega-2)t +[1/4(\omega+2)]\sin2\omega_0[\cos(\omega+2)t-1] -[1-4(\omega-2)]\sin2\varphi_0[\cos(\omega-2)t-1]. \quad (3.1)$$

It is seen that if $\omega\neq2$, θ_1 is a periodic function with period 2π (if ω is integer) containing harmonics. It is to be noted that the third term on the right side of Eq. (3.1) at the limit $\omega=2$ becomes the secular term appearing in Eq. (2.11). In fact, it is sufficient to set $\omega-2=x$ and to pass to the limit $(\sin xt/x)_{x=0}=t$, which gives that term. Likewise, one finds that the last term in (3.1) is zero at the limit $\omega=2$. In a similar manner one ascertains that the expression for $P_1(t)$ does not contain secular terms; namely,

$$P_1(t)=-\sin2\varphi_0\{[1/2(\omega+2)]\sin(\omega+2)t +[1/2(\omega-2)]\sin(\omega-2)t\} -\cos2\varphi_0\{[1/2(\omega+2)][\cos(\omega+2)t-1] -[1/2(\omega-2)][\cos(\omega-2)t-1]\}, \quad (3.2)$$

inasmuch as all terms are periodic if $\omega\neq2$. One finds that the second term on the right of Eq. (3.2) degenerates into a secular term at the limit $\omega=2$ and that P_2 does not contain secular terms, while θ_2 does. This means that p.e. exists only in still higher orders and for that reason is negligible.

4. NONLINEAR CASE

The great majority of cases encountered in applications are of the nonlinear type. The Melde-Rayleigh experiment is an example of this type. In fact, the stationary amplitude in this case is reached owing to the nonlinear elasticity of the vibrating wire for larger amplitudes. Moreover, in all cases there is a dissipation of energy in one form or another.

The perturbation procedure outlined in the preceding sections gives a convenient tool for the investigation of the nonlinear range of the phenomenon of p.e. to which the classical methods of d.e. with periodic coefficients do not apply. For that reason, instead of Eq. (2.1) we shall investigate now a d.e. of the form,

$$\ddot{x}+p\dot{x}+f(x)+\gamma x\cos2t=0,$$

where, as is frequently done in applications, the nonlinear term $f(x)$ may be assumed to be of the form $x+\epsilon x^3$; that is, we will consider the equation

$$\ddot{x}+p\dot{x}+(1+\gamma\cos2t)x+\epsilon x^3=0. \quad (4.1)$$

Moreover, we will assume that ϵ and p are small numbers of the same order as γ. The form (4.1) results from keeping only the terms of the first order and assuming $\omega=2$ as previously.

The perturbation procedure yields the following differential equations:

$$dP_1/dt=-2B\sin^2\theta_0-\cos2t\sin2\theta_0 -\tfrac{1}{2}A\rho_0(\sin2\theta_0+\tfrac{1}{2}\sin4\theta_0), \quad (4.2)$$

$$d\theta_1/dt=-\tfrac{1}{2}B\sin2\theta_0-\cos2t\cos^2\theta_0-A\rho\cos^4\theta_0, \quad (4.3)$$

where $A=\epsilon/\gamma$, $B=p/\gamma$, and $\theta_0=\varphi_0-t$ as before.

The integration yields

$$P_1=-\tfrac{1}{2}(\sin2\varphi_0+2B)t+\text{periodic terms}, \quad (4.4)$$

$$\theta_1=-\tfrac{1}{4}(\cos2\varphi_0+\tfrac{3}{2}A\rho_0)t+\text{periodic terms}. \quad (4.5)$$

After one period 2π the quantities P_1 and θ_1 undergo the variations,

$$\Delta P=-\pi\gamma(\sin2\varphi_0+2B)+0(\gamma^2);$$
$$\Delta\varphi=-\tfrac{1}{2}\pi\gamma(\cos2\varphi_0+\tfrac{3}{2}A\rho_0)+0(\gamma^2). \quad (4.6)$$

Setting $2\pi\gamma=\Delta t$ and considering Δt as dt in the problem involving the study of motion in the course of many periods 2π, one can replace the difference equations (4.6) by the following differential equations describing the behavior of the system in the long run

$$d\rho/dt=-\tfrac{1}{2}\rho(\sin2\varphi+2B), \quad (4.7)$$

$$d\varphi/dt=-\tfrac{1}{4}(\cos2\varphi+\tfrac{3}{2}A\rho). \quad (4.8)$$

The singular point of this system is given by equations,

$$\sin2\varphi_0=-2B;\quad \cos2\varphi_0=-\tfrac{3}{2}A\rho_0,$$

which gives

$$\rho_0=(\tfrac{2}{3}A)(1-4B^2)^{\frac{1}{2}}. \quad (4.9)$$

In order to investigate the stability of the singular point, we form the variational equations of Eqs. (4.7) and (4.8), viz.:

$$d\delta\rho/dt=-\rho_0\cos2\varphi_0\cdot\delta\varphi;\quad d\delta\varphi/dt=-\tfrac{3}{8}A\delta\rho +\tfrac{1}{2}\sin2\varphi_0\delta\varphi. \quad (4.10)$$

The characteristic equation of (4.10) is

$$S^2-\tfrac{1}{2}\sin2\varphi_0S-\tfrac{3}{8}A\rho_0\cos2\varphi_0=0. \quad (4.11)$$

Replacing the coordinates of the singular point ρ_0, φ_0 by their values gives

$$S^2+BS+\tfrac{1}{4}(1-4B^2)=0. \quad (4.11a)$$

The singular point is a saddle point if $1-4B^2<0$, that is, if $\gamma<2p$. If $1-4B^2>0$, one has either a nodal point if $\gamma<(5)^{\frac{1}{2}}p$, or a focal point if $\gamma>(5)^{\frac{1}{2}}p$, these singularities being stable.

From Eq. (4.9) one notes that the condition of reality of the stationary amplitude ρ_0 is $\gamma>2p$, that is, the same as for the absence of the saddle point, which is obvious, since no closed periodic trajectory can exist enclosing a singular point with index $j=-1$.

For a conservative system ($B=0$), Eq. (4.11a) becomes $S^2+\frac{1}{4}=0$, that is, the singular point is a vortex

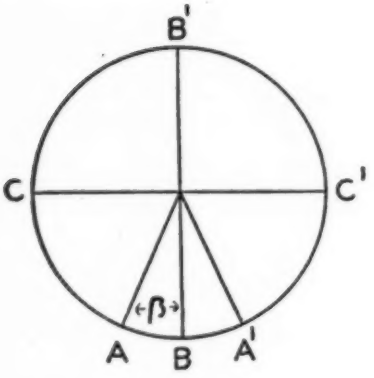


FIG. 1. Sector of self-excitation.

point. In this case from Eq. (4.9)

$$\rho_0 = 2\gamma/3\epsilon. \tag{4.12}$$

We conclude, therefore, that a conservative nonlinear system excited parametrically is generally unstable in the sense in which the word "stability" is understood in application. In other words, if a perturbation deviates the system from its stationary condition, this effect will not be reduced in time, since the representative point in the (ρ, φ) plane will describe indefinitely a closed curve containing the vortex point in its interior. In the (ρ, θ) plane this circumstance results in parasitic amplitude and phase modulations. Since the disturbances are erratic, so will be these modulations, although on *the average* there will be still a stability (in the sense of Liapounoff) in that the motion will remain in a bounded region of the phase plane. From this standpoint the presence of a dissipation of energy is beneficial in that it smoothes out these parasitic phenomena.

5. CONDITIONS OF SELF-EXCITATION OF NONLINEAR SYSTEMS

In the case of linear systems (2.4), (2.5) it was possible to eliminate the variable ρ on the right-hand side of (2.4) by introducing an auxiliary variable $P(t)$ and to thus reduce the system to a relatively simple form. For a nonlinear system this is not the case, and it is necessary to limit the analysis to a qualitative discussion. A difficulty presents itself if one attempts to set $\rho=0$ in order to determine the conditions of self-excitation from rest. Since ρ is the energy content of the system, it is clear that the condition $\rho=0$ entails also that all other dynamical variables (currents, voltages, velocities, etc.) are also zero, in which case the concept of the *phase* φ loses its physical significance.

In reality there always exists a small initial energy content in any physical system. Thus, for instance, if the system is electrical, stray charges on a condenser account for this initial electrostatic energy ρ_1. As to the phase φ, it is to be assumed as entirely arbitrary. For that reason it is necessary to investigate the starting of the phenomenon for *any value* of the initial phase φ_1.

From Eq. (4.7) it follows that $d\rho/dt>0$ if $(\sin 2\varphi + 2B)<0$. Assuming first $0<2B<1$, this condition is fulfilled if $\sin 2\varphi<0$ and $|\sin 2\varphi|>2B$. This defines a sector AOA' (Fig. 1) which we may call the *sector of self-excitation* (s.s.e. for short). If 2φ is in s.s.e., ρ increases. If 2φ is outside this sector, the small initial ρ_1 will be further reduced. The semi-angle β determining the s.s.e. is given by the relation $\cos\beta=2B$.

For reasons which will appear later, we shall consider the boundaries OA and OA' of the s.s.e. as belonging to it. For a nondissipative system the s.s.e. becomes the lower half-circle CBC'. When $\gamma=2p$ (which is the threshold between the region of saddle points and that of nodal points), the s.s.e. shrinks to the line OB.

We shall investigate four different cases when the initial phase $2\varphi_0$ is situated in the sectors BOA', AOB, $B'OA'$, and $B'OA$.

In BOA' $\cos 2\varphi_1>0$ and, therefore, $d\varphi/dt<0$. The phase angle will move clockwise and will enter the sector AOB. Since in this sector $\cos 2\varphi<0$, the two terms on the right side of Eq. (4.8) are of opposite signs, so that $d\varphi/dt=0$ when $|\cos 2\varphi| = \frac{3}{2}A\rho$. In AOB ρ increases while the phase, for which $d\varphi/dt=0$, shifts gradually toward OA. During this shift $d\rho/dt$ decreases and becomes zero at OA, where $d\varphi/dt$ also becomes zero by Eq. (4.8). The phase angle $2\varphi_0$ of the singular point is thus OA.

If the initial phase $2\varphi_1$ is in $B'OA'$, $d\varphi/dt<0$ and, since in this sector $d\rho/dt<0$, the small initial value ρ_1 is still further reduced, and we can neglect the second term on the right of Eq. (4.8). Under these conditions, the phase $2\varphi_0$ will enter the s.s.e. through the boundary OA' and will reach the stable phase OA in the manner which we have just investigated.

If $2\varphi_1$ is in $B'OA$, $d\varphi/dt>0$ so that the phase will enter the s.s.e. through its boundary OA. This time OA will not be a position of equilibrium for 2φ on account of the fact that $\rho\cong 0$ during this passage through OA. Therefore, the phase 2φ will enter AOB, where ρ will increase to a value for which $d\varphi/dt=0$. Since in AOB ρ is increasing, this position of equilibrium for ρ will gradually shift itself toward OA, where both positions of equilibrium for ρ and for φ will be simultaneously attained as was previously shown.

Summing up, from any initial phase $2\varphi_1$, the ultimate stable phase $2\varphi_0$ will be OA; and when the phase settles on this value, the variable ρ also reaches a stable equilibrium which corresponds thus to the singular point of the system (4.7) and (4.8).

It remains to investigate the case when $B>\frac{1}{2}$ (i.e., when $\gamma<2p$). From Eq. (4.7) it follows that $d\rho/dt<0$ for any 2φ. Since ρ, originally small, decreases continuously, from Eq. (4.8) it follows that $\cos 2\varphi$ determines the stability of the phase. By a simple argument one finds that $2\varphi_0=\pi/2$ is an unstable phase and $2\varphi_0=3\pi/2$ is a stable phase. This case corresponds at the same time to the condition for the existence of a saddle point.

We are thus led to identify the existence of a saddle

point of Eqs. (4.7) and (4.8) with the loss of the initial energy of the system or, using a different language, when the parameter variation is too small to be able to "cross the threshold" of parametric excitation. It is useful to mention that no such threshold exists in the case of a nondissipative system ($p=B=0$). In fact, the condition $\gamma<2p$ in this case would be $\gamma<0$, which is impossible, since γ is an essentially positive quantity. We saw that the only possible equilibrium in this case is that of a vortex point.

6. EXPERIMENTAL EVIDENCE

From the preceding it follows that the parameter variation is the *cause* of the phenomenon and the cumulative excitation of the system is its *effect*. Hence, any perturbation in the former inevitably reacts on the latter. Moreover, the phenomenon is possible on account of the stability of the phase. If, therefore, the latter is disturbed, the disturbance also reacts on the amplitude. One can easily demonstrate this effect by disturbing the phase of the parameter variation. In this manner the old phase of motion ceases to be a stable phase φ_0'' relatively to the new phase of the parameter, so that a period of readjustment of the phase of motion to the new phase of the parameter takes place. During this transient period, investigated in Sec. 2, the energy injection into the system is also modified, resulting in a certain "dip" of the amplitude as compared to its value prior to this sudden change. After the phase adjustment has taken place, the amplitude resumes its former value.

The oscillogram of Fig. 2 shows this phenomenon. The vibrating wire (as in Lord Rayleigh's experiment) was used in this experiment as a coupling capacitor, the other side being a small aluminum plate under the wire. A 200-kilocycle carrier signal passing through this capacitor was thus modulated by the wire vibration. The modulated signal was amplified through several tuned stages and detected. The detected signal was fed both to the vertical amplifier of a cathode ray oscilloscope for a visual observation and through a direct current amplifier to a Brush BL202 recorder. The phase of the exciting 60-cycle current in the magnet was changed once a second by switching in and out of the circuit an auxiliary series condenser-resistor combination. The phase change so produced was about 89° with the voltage across the magnet coil remaining practically constant. The record as given in Fig. 3 shows the normal decay of the wire with the excitation removed.

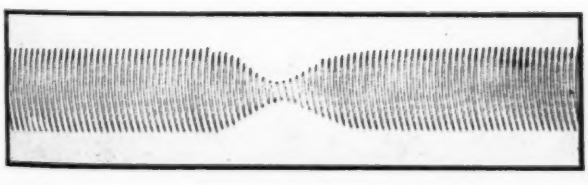


FIG. 2. Effect of the variation of the phase on the amplitude.

FIG. 3. Decay of an oscillation with the excitation removed

The author is indebted to Mr. M. Blumberg for this experiment.

7. PARAMETRIC EXCITATION BY A RECTANGULAR RIPPLE

Instead of a sinusoidal variation of the parameter, the p.e. can be obtained also by means of a discontinuous variation of the parameter. We shall consider the simplest case of a rectangular ripple where during fixed time intervals the parameter remains in its maximum or minimum values. This case can be analyzed in an elementary way and already leads to all phenomena connected with parametric excitation.

The differential equation of the rectangular ripple excitation is of the so-called Hill-Meissner type and can be represented by the two elementary equations

$$\ddot{x}+\alpha_1^2 x=0 \quad \ddot{x}+\alpha_2^2 x=0 \tag{7.1}$$

replacing each other after fixed prescribed time intervals. The solution is required to vary continuously in time.

The integral curves of the Eqs. (7.1) are ellipses

$$(x^2/a^2)+(y^2/b^2)=1, \tag{7.2}$$

where $\frac{1}{2}b^2=h$ measures the energy of the system at the moment considered and $a^2=2h/\alpha_i^2$. Thus, the representative point (x, y) in phase space will describe a path which is composed of elliptic arcs belonging to two families of concentric ellipses with axis ratio α_1 or α_2, respectively.

In order to deal with a particularly simple and instructive situation, let us assume that the intervals of the rectangular ripple are chosen in such a way that the parameter α_1 is replaced by the parameter α_2 when the representative point crosses the axis $x=0$ and that α_2 is again replaced by α_1 whenever the point crosses the axis $y=0$. The point (x, y) needs the time $\pi[(1/\alpha_1)+(1/\alpha_2)]$ in order to describe an angle 2π around the origin, and in this time the parameters α_1 and α_2 hold twice during time intervals $\pi/2\alpha_1$ and $\pi/2\alpha_2$, respectively. This model bears some analogy to the case $\omega=2$ of continuous parameter variation and should be compared with it.

Let us study now the evolution of the energy of the system if the representative point describes one turn around the origin. It is clear that if we fit two ellipses of the type (7.2) continuously together at the axis $x=0$, we will have to require that both have the same value b, while if we put them continuously together at the axis $y=0$, we will have to require the same value of a for both. Since $b=2h$ is closely related to the energy of the system, we recognize that no energy change

takes place if the representative point crosses the axis $x=0$. This is physically evident, since at this moment the whole energy of the system is kinetic and none of it is potential; hence, a sudden change of the constant of potential energy can have no influence on the energy content of the system. If, however, the axis $y=0$ is crossed the condition that a be unchanged leads to the condition

$$h/\alpha_2^2=h'/\alpha_1^2 \tag{7.3}$$

if h is the energy content before the crossing and h' after it. Thus, in one turn around the origin the original energy is to be increased (or decreased) in the ratio

$$h_2/h_1=(\alpha_1/\alpha_2)^4. \tag{7.4}$$

If $\alpha_1>\alpha_2$, the system will absorb energy in each turn; but if $\alpha_1<\alpha_2$, the energy will be drained away from it. This reduction of energy has in all respects the character of a *damping*, but this damping is effected not by a *dissipation* of energy but by making the system do external work. By analogy with the term "parametric generator" one could designate this performance as "parametric damper."

If $\alpha_1>\alpha_2$ are given, the described type of parametric variation leads to a continuous increase of energy of the system. Let now this process of variation be shifted by a quarter phase against the phase of the motion; i.e., change from α_1 to α_2 if the $y=0$ axis is crossed and from α_2 to α_1 if the $x=0$ line is passed. The roles of α_1 and α_2 are interchanged and instead of parametric excitation we obtain parametric damping. Analogous calculations can easily be carried out for any phase difference ψ between the parameter change and the phase of the motion. The phase difference ψ attained by the system for given rectangular ripple is determined by stability considerations.

The analysis contained in this section does not involve any question of stability. On the other hand, from a direct discussion of the differential equation (Sec. 2) it was found that there exists a stable phase at which the phenomenon stabilizes itself; this phase corresponds also to the maximum absorption of energy. Hence, this is the only case which is accessible to the experiment, all other phases being ruled out by the condition of stability. This applies particularly to the range in which the parametric damping occurs.

As far as known, there exists no experimental evidence regarding the existence of parametric damping, although on the basis of the preceding argument it seems plausible to assume that it exists if proper experimental means are provided. One can, for example, envision the following arrangement: assume that in the Melde-Rayleigh experiment we produce sustained lateral oscillations of the wire by some means. By deriving a voltage from this oscillation and shifting it by means of a phase shifting network by an appropriate phase angle, one may expect that one can reach a point for a definite value of the phase at which the operation of the parameter, instead of exciting the lateral oscillations, will, on the contrary, extinguish them.

A "feed-back" arrangement of this kind will *impose* an arbitrary phase to the parameter instead of operating with the *free* phase, so to speak, which adjusts itself automatically to the stable point by virtue of the stability condition mentioned in connection with Eqs. (2.16) and (2.17).

In conclusion, it is useful to note that it seems plausible to admit that the conclusions obtained on the basis of the Hill-Meissner equation are applicable also to the Mathieu equation, as was pointed out in the past, although there exists no formal justification of this point as far as is known. One observes this indirectly from the fact that the formulas (2.18, Mathieu) and (7.7, Hill-Meissner) for small values of γ give a fairly consistent result especially because (2.18) relates only to the first approximation.

The author wishes to express his gratitute to Professor M. Schiffer for valuable discussions of this matter.

This work was carried out under the authority of the ONR. The permission to publish this paper is greatly appreciated.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Effective Earth's Radius for Radiowave Propagation Beyond the Horizon

WILLIAM MILLER
Central Radio Propagation Laboratory, National Bureau of Standards, Washington, D. C.
(Received June 12, 1950)

The geometrical optical arguments underlying the equivalent earth's radius approximation are extended to non-horizontal rays. The appropriate wave equation for a non-uniform, but spherically symmetric, region is derived in a natural way. The notion of a "primary field" with the form e^{ikR}/R at the source is dropped in favor of a solution to the wave equation in terms of a Green's Function. The formal solution thus obtained is seen to contain the known solutions for a uniform medium. For the non-uniform case, the solutions to the radial equation are found by a technique due to Langer. These functions account for the index variation near the earth without making unwarranted assumptions about the behavior at greater heights. The formal series obtained is summed by the Watson technique. The first term in this series (lowest mode) alone determines the field at large distances and indicates that one may account for standard atmospheric refraction by using an effective earth's radius.

I. INTRODUCTION

THE increased use of frequencies above 50 Mc/sec. for radio communication has focused attention on the propagational effects of tropospheric refraction. Although this refraction exists at lower frequencies, it is there overshadowed by ionospheric reflection and is thus not a prime factor in determining field intensities.

To calculate tropospheric fields, one must consider the problem of an oscillating doublet (infinitesimal dipole) above a sphere which is surrounded by a non-uniform medium. The free space wave-length, λ_0, of the radiation and the radius, a, of the sphere involved are such that $a/\lambda_0 > 10^6$. Formal solutions to this problem, with the atmosphere assumed homogeneous, were available as early as 1900.[1] However, the expressions for the field vectors were in the form of a slowly converging series of integral order spherical harmonics; it was necessary to calculate at least a/λ_0 terms to get numerical results. This difficulty was overcome by Watson[2] who first transformed the series into a contour integral, and then into a rapidly converging[3] series of spherical harmonics of complex order. This work was later extended by Van der Pol and Bremmer.[4]

To account for refraction by the troposphere, Schelleng, Burrows, and Ferrell[5] applied geometrical optical ideas. Attention was concentrated on a standard atmosphere which may be defined as one in which the index of refraction is a linear function of height near the earth, and approaches 1 monotonically at greater heights. A "ray" then travels in an approximately circular path, with a radius of curvature several times that of the earth for normal lapse rates of the refractive index. It is then argued that, since the same distance *vs.* height relationship can be maintained between the ray and earth by drawing the ray in a straight line, and enlarging the radius of the earth so as to maintain the same relative curvature between ray and earth, the radiation from a source in a uniform atmosphere above an earth with the appropriate effective radius should yield the same field strength as a source in a standard atmosphere above the actual earth. This argument will be extended somewhat in the next section.

A treatment of tropospheric refraction based on the "*WKB*" or phase-integral[6] method was given by Eckersley and Millington.[7] The field vectors are assumed to satisfy the wave equation[8]

$$(\nabla^2+k^2)\psi=0, \tag{1}$$

where the propagation constant, $k(\equiv 2\pi/\lambda)$ is a function of r only. The field is then expressed in the form

$$\psi=\sum_\nu s_\nu(r)P_\nu(-\cos\theta), \tag{2}$$

where the ν's are as yet undetermined, and the terms are to be combined so as to give a singularity at the source. Substitution of (2) into (1) indicates that $s_\nu(r)$ is simply related to the solution of

$$\frac{d^2T_\nu}{d\rho^2}+\left\{\left(\frac{k(\rho)}{k_0}\right)^2-\frac{\nu(\nu+1)}{\rho^2}\right\}T_\nu=0, \tag{3}$$

where $\rho=2\pi r/\lambda_0$. This may be likened to the Schroedinger wave equation of quantum mechanics. For $\rho\to\infty$, $k/k_0\to 1$, $\nu(\nu+1)/\rho^2\to 0$, and T_ν is a simple oscillatory function. For $\rho<k_0a$, $k^2(\rho)/k_0^2$ is large and complex giving a rapidly oscillating, attenuated wave. The numbers ν are then chosen so that the *WKB* solution to (3) for intermediate $\rho(>k_0a)$ having the form $\exp\{\pm 2\pi i\int[(k/k_0)^2-\nu(\nu+1)/\rho^2]^{\frac{1}{2}}d\rho\}$ satisfies the appropriate boundary conditions on T_ν and its derivatives, which lead to a condition on the exponent, the phase

[1] For a summary of the early work and references see A. E. H. Love, Phil. Trans. Roy. Soc. (London) (A) **216**, 105–131 (1915).
[2] G. N. Watson, Proc. Roy. Soc. (London) (A) **95**, 83–99 (1918).
[3] This method is not suitable for calculating field strengths near the transmitter. See K. A. Norton, Proc. Inst. Radio Engrs. **29**, 623–639 (December, 1941).
[4] B. Van der Pol and H. Bremmer, Phil. Mag. **24**, 141–176 (July, 1937); **24**, 825–864 (November, 1937); **25**, 817–834 (June, 1938).
[5] Schelleng, Burrows, and Ferrell, Proc. Inst. Radio Engrs. **21**, 427–463 (March 1933).
[6] See, e.g. E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill, Book Company, Inc., New York, 1937).
[7] T. L. Eckersley and G. Millington, Phil. Trans. Roy. Soc. (London) (A) **237**, 273–309 (1938).
[8] This point is discussed in Section III.

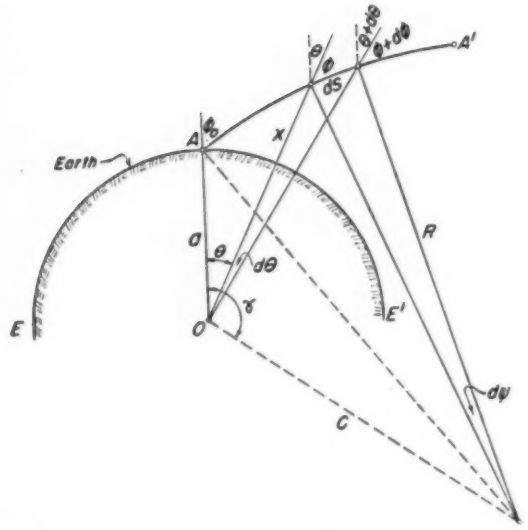


FIG. 1. Ray diagram for propagation in the earth's atmosphere. The curve AA′ represents the path of the ray.

integral condition. The conditions at $\rho=k_0a$ were satisfied by using the known results for reflection of a plane wave at a plane interface, and to this extent, the calculation was not completely self-contained. By letting $k^2/k_0^2=1-\eta+(\epsilon a^2/r^2)$ and then comparing the results with those for $\eta=\epsilon=0$, Eckersley and Millington find the effective earth's radius approximation to be valid.

The radiation from an infinitely long wire parallel to the axis of an infinitely long conducting cylinder, surrounded by a medium with index varying as a/r was examined by Freehafer,[9] who concluded that an effective radius approximation was valid in that case. It is interesting to note that the assumed index dependence specifies an infinite velocity for $r\to\infty$.

Recent measurements indicate that field strengths deep in the diffraction region (200–400 Mi) are several orders of magnitude larger than values calculated on an equivalent earth basis. Since these fields are large all the time, one must conclude that either the atmospheric conditions are never standard or the equivalent earth's radius approximation is not valid in the diffraction region. In order to definitely eliminate the latter alternative, a self-contained wave solution is considered here, since the ray-tracing arguments, which were originally the justification for the equivalent earth's radius approximation, may not be used to calculate diffracted fields. First, however, the ray-tracing arguments will be extended to include non-horizontal rays.

II. GEOMETRIC-OPTICAL CONSIDERATIONS

If a transmitter and receiver are located at points A and A' above the earth, then the path of a "ray" travelling between them is determined by Fermat's principle,[10] which states that for the actual path the time of traversal be an extreme, i.e.,

$$\delta\int_{r_A}^{r_{A'}} n(r)(1+r^2(d\theta/dr)^2)^{\frac{1}{2}}dr=0, \tag{4}$$

where $n(=k/k_0)$ is the index of refraction. Applying the appropriate Euler condition[11] gives

$$d/dr\{[\partial/\partial(d\theta/dr)][n(r)(1+r^2(d\theta/dr)^2)^{\frac{1}{2}}]\}=0 \tag{5}$$

or

$$\frac{n(r)r^2d\theta}{(dr^2+r^2d\theta^2)^{\frac{1}{2}}}=nr\sin\varphi=\text{constant}, \tag{6}$$

where φ is the angle made by the tangent to the ray and the radius vector (see Fig. 1). The curvature, $1/R$, of the ray is given by:

$$\left|\frac{1}{R}\right|=\left|\frac{d\psi}{ds}\right|=\left|\frac{d\varphi+d\theta}{ds}\right|. \tag{7}$$

Differentiating (6), we find:

$$\begin{aligned} d\varphi&=-[(dn/n)+(dr/r)]\tan\varphi \\ &=-\tan\varphi(dn/n)-d\theta. \end{aligned} \tag{8}$$

Thus

$$|1/R|=|1/n(dn/dr)\sin\varphi|. \tag{9}$$

To establish an equivalent earth's radius approximation the ray is replaced by a straight line, i.e. given zero curvature, and the radius of the earth is adjusted so as to maintain the same distance height relationship. This is usually done by maintaining the same relative curvature between straight ray and enlarged earth as between refracted ray and actual earth. a_{eff} is then given by

$$1/a_{\text{eff}}=1/a-1/R \tag{10}$$

which, *for almost horizontal rays*, i.e., $\sin\varphi\cong 1$, takes the form

$$1/a_{\text{eff}}=1/a+1/n(dn/dr). \tag{11}$$

If the procedure (10) were the correct one for maintaining the distance height relationship between ray and earth, (9) would indicate that a_{eff} would vary with the angle of departure of the ray. We may investigate this point by considering the circle EE', and the arc AA' (Fig. 1). Restricting θ and x to small values will keep the curvature along the path essentially constant, and we may write

$$\begin{aligned} &c\cos\gamma=a-R\sin\varphi_0, \\ &c\sin\gamma=R\cos\varphi_0, \\ &R^2=c^2+(a+x)^2-2c(a+x)\cos(\gamma-\theta). \end{aligned} \tag{12}$$

[9] J. E. Freehafer, M.I.T. Rad. Lab. Report 447.

[10] See e.g., J. C. Slater and N. H. Frank, *Introduction to Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1933), pp. 339–342.

[11] See, e.g., H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand, Inc., New York, 1943), Chapter VI.

Solving for x gives:

$$x \cong a\left\{\frac{\sin\varphi_0}{a/R(1-\cos\theta)+\sin(\varphi_0-\theta)}-1\right\}. \quad (13)$$

Now, if $R\to\infty$, $a\to a_{eff}$, and the coordinates of this straight ray are θ', x' we find

$$x' \cong a_{eff}\left\{\frac{\sin\varphi_0}{\sin(\varphi_0-\theta')}-1\right\}. \quad (14)$$

Taking $\theta' a_{eff}=\theta a$ and requiring that $x'=x$, we find after some reduction that

$$\frac{1}{a_{eff}}=\frac{1}{a}-\frac{1}{R\sin\varphi_0(1+2\cot^2\varphi_0)} \quad (15)$$

which indicates that (10) is not applicable for non-horizontal rays. Using (9) to find R, the effective earth's radius is given by

$$\frac{1}{a_{eff}}=\frac{1}{a}-\frac{1}{n}\frac{dn}{dr}\frac{\sin\varphi}{\sin\varphi_0}\frac{1}{(1+2\cot^2\varphi_0)}. \quad (16)$$

The effective radius concept will be a useful one when the value of a_{eff}, as given by (16), is nearly constant over the path of the ray. This is most nearly true when $\ln n(\approx n-1)$, is a linear function of r, and $\varphi\approx\pi$, so that $\sin\varphi$ is slowly varying. Assuming these conditions to be satisfied, one then argues that if the field intensity is assumed proportional to the number of rays per unit area, with the rays symmetrically distributed at the source, then the requirements $x=x'$ and $\theta a=\theta' a_{eff}$ would be sufficient to guarantee that the straight ray enlarged earth model would predict the same primary field intensity distribution as the curved ray picture. Schulkin[12] has made a detailed study of tropospheric conditions at various localities and has indicated the applicability of the approximation in specific cases.

Any discussion of the diffracted field (i.e., beyond line of sight) or effects of reflection from a non-perfectly conducting earth on the basis of this ray picture would be quite difficult. Thus, for a more rigorous treatment of the validity of this effective radius concept, one must resort to a wave treatment.

III. WAVE EQUATION APPROPRIATE TO NON-UNIFORM MEDIUM

We start with Maxwell's equations for the field vectors[13]

$$\begin{aligned}&\text{(I)}\quad \nabla\times\mathbf{H}=\epsilon'\frac{\partial\mathbf{E}}{\partial t}+\mathbf{J}_0, \qquad \text{(III)}\quad \nabla\cdot\mathbf{B}=0,\\ &\text{(II)}\quad \nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}, \qquad \text{(IV)}\quad \nabla\cdot\mathbf{D}=\rho,\\ &\mathbf{B}=\mu\mathbf{H}, \quad \mathbf{D}=\epsilon'\mathbf{E},\end{aligned} \quad (17)$$

where ϵ' is the complex dielectric constant, so that $\epsilon'(\partial\mathbf{E}/\partial t)$ includes both displacement and conduction currents, and $\mathbf{J}_0$ represents "forced" currents such as those in a driven antenna. A time dependence $e^{-i\omega t}$ has been assumed. Considering, first, that ϵ' and μ are not functions of position, the solution to (17) is obtained as follows:

Taking

$$\mathbf{B}=\nabla\times\mathbf{A} \quad (18)$$

satisfies (III), and when substituted into (I) gives

$$\frac{\partial\mathbf{E}}{\partial t}=\frac{1}{\epsilon'\mu}\nabla\times\nabla\times\mathbf{A}-\frac{\mathbf{J}_0}{\epsilon'}. \quad (19)$$

Thus far, $\mathbf{A}$ is an arbitrary vector field, but a condition on $\mathbf{A}$ may be found by inserting (18) and (19) into (17 II), yielding

$$\nabla\times\left(\frac{1}{\epsilon'\mu}\nabla\times\nabla\times\mathbf{A}-\frac{\mathbf{J}_0}{\epsilon'}+\frac{\partial^2\mathbf{A}}{\partial t^2}\right)=0. \quad (20)$$

This will be satisfied by setting

$$\frac{1}{\epsilon'\mu}\nabla\times\nabla\times\mathbf{A}-\frac{\mathbf{J}_0}{\epsilon'}+\frac{\partial^2\mathbf{A}}{\partial t^2}=-\nabla\frac{\partial\phi}{\partial t}, \quad (21)$$

where $\partial\phi/\partial t$ is an arbitrary function of position, corresponding to the fact that (18) and (19) are invariant to additions of arbitrary gradients to $\mathbf{A}$. That Eq. (IV) is satisfied may be seen by taking the divergence of (I) and applying the continuity equation for charge.

Various equations for $\mathbf{A}$ may be obtained depending on the choice of $\partial\phi/\partial t$. The most commonly used expression for $\partial\phi/\partial t$ is given by the *Lorentz relation*[14]

$$\mu\epsilon'\frac{\partial\phi}{\partial t}=-\nabla\cdot\mathbf{A}. \quad (22)$$

For constant ϵ' and μ, this will reduce (21) to

$$(\nabla^2+k^2)\mathbf{A}=-\mu\mathbf{J}_0, \quad (23)$$

where the time dependence $e^{-i\omega t}$ has been assumed and $k^2=\omega^2\mu\epsilon'$. This implies a corresponding scalar equation for each of the rectangular components of $\mathbf{A}$ and $\mathbf{J}_0$, e.g.,

$$(\nabla^2+k^2)A_x=-\mu J_{0x}, \quad (24)$$

[12] M. Schulkin, "Average Radio Ray Refraction in the Lower Atmosphere," Report CRPL 2-2, Aug. 1947, Natl. Bur. Standards, Washington, D. C.

[13] For definitions of these quantities and the M.K.S. units used here see J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941).

[14] See W. Heitler, *The Quantum Theory of Radiation*, second edition (Oxford University Press, London, 1944), Chapter I.

so that if $\mathbf{J}_0$ has components along one cartesian axis only, it is necessary to solve only one inhomogeneous equation of the form (24). However, if $\mathbf{J}_0$ does not point in the same direction everywhere, Eq. (23) is no longer convenient, since it would be necessary to solve two or three inhomogeneous equations.

One case especially worthy of note involves $\mathbf{J}_0$ radial. Here it would be desirable to choose $\partial\phi/\partial t$ so that (21) contains terms with radial components only.[15] If one tries a radial $\mathbf{A}$, i.e., $\mathbf{A}=A\mathbf{r}_1$, ($\mathbf{r}_1$ is a unit vector), abandons the Lorentz relation (22) and uses instead

$$\mu\epsilon'(\partial\phi/\partial t)=-\partial A/\partial r \tag{25}$$

one finds that (21) becomes

$$\frac{\partial^2 A}{\partial r^2}+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial A}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2 A}{\partial\varphi^2}$$
$$-\mu\epsilon'\frac{\partial^2 A}{\partial t^2}=-J_0\mu, \quad (\mathbf{J}_0=J_0\mathbf{r}_1), \tag{26}$$

A may be readily related to a solution of the wave equation, for if one defines

$$\psi\equiv A/r \tag{27}$$

then it follows that

$$(\nabla^2+k^2)\psi=-\frac{J_0\mu}{r}. \tag{28}$$

Frequently, there is introduced a *Hertz potential* which is defined in the rectangular case by $\mathbf{A}=\partial\Pi/\partial t$, and in the radial case by $\mathbf{A}=\mathbf{r}(\partial\Pi/\partial t)r$, so that

$$(\nabla^2+k^2)\Pi=-\mu\mathbf{P}_0 \tag{23'}$$

and

$$(\nabla^2+k^2)\Pi_r=\frac{\mu P_{0r}}{r}, \tag{28'}$$

where $\mathbf{P}_0$ is defined as the free charge polarization, ($\mathbf{J}_0=\partial\mathbf{P}_0/\partial t$). If ϵ' and μ are functions of position, then the relations (22) and (25) no longer lead to the wave Eqs. (23) and (28). We shall be particularly concerned with the radial vector potential, constant μ, and an ϵ' which is a function of r only, so that if (25) is now inserted into (21) and we again set $A=\psi r$, we find

$$\frac{\partial^2\psi}{\partial r^2}+\frac{\partial\psi}{\partial r}\left\{\frac{2}{r}+\epsilon'\mu\frac{d}{dr}\left(\frac{1}{\epsilon'\mu}\right)\right\}+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)$$
$$+\frac{1}{r^2\sin^2\varphi}\frac{\partial^2\psi}{\partial\varphi^2}+\psi\left[\omega^2\mu\epsilon'+\frac{\epsilon'\mu}{r}\frac{d}{dr}\left(\frac{1}{\epsilon'\mu}\right)\right]=-\frac{J_0\mu}{r}, \tag{29}$$

which would reduce to (28) for ϵ', μ constant. To cast (29) into the form of the wave equation one might proceed by: (a) Introducing a new radial coordinate $r'=r'(r,\epsilon')$; (b) Introducing modified Hertz potential through $\psi=(\partial\Pi/\partial t)g(r)$. (a) leads to two simultaneous differential equations for r', with no common solution for arbitrary ϵ'. Substituting (b) into (29) leads to

$$\frac{\partial^2\Pi}{\partial r^2}+\frac{\partial\Pi}{\partial r}\left\{\frac{2}{g}\frac{dg}{dr}+\frac{2}{r}+k^2\frac{d}{dr}\left(\frac{1}{k^2}\right)\right\}$$
$$+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Pi}{\partial\theta}\right)+\frac{1}{r^2\sin2\theta}\frac{\partial^2\Pi}{\partial\varphi^2}$$
$$+\Pi\left\{k^2+\frac{k^2}{g}\frac{dg}{dr}\frac{d}{dr}\left(\frac{1}{k^2}\right)+\frac{1}{g}\frac{d^2g}{dr^2}\right.$$
$$\left.+\frac{k^2}{r}\frac{d}{dr}\left(\frac{1}{k^2}\right)+\frac{2}{gr}\frac{dg}{dr}\right\}=-\frac{1}{g}\frac{P_{0r}\mu}{r}. \tag{30}$$

This will have the form of the wave equation if

$$\frac{2}{g}\frac{dg}{dr}=-k^2\frac{d}{dr}\left(\frac{1}{k^2}\right) \tag{31}$$

i.e.

$$g=k$$

and a propagation constant k^* is used where

$$k^*=k^2-\frac{2}{k^2}\left(\frac{dk}{dr}\right)^2+\frac{1}{k}\frac{d^2k}{dr^2} \tag{32}$$

which is seen to agree with the quantity k^* used by Bremmer.[16] Figure 3 is a typical curve[12] of k/k_0 and it is seen that $|(k^{*2}-k^2)/k_0^2|<220\cdot10^{-9}$, so that k^* and k may be considered identical.

To recapitulate, we find $\mathbf{E}$ and $\mathbf{B}$ from (18) and (19) where

$$\mathbf{A}=\psi\mathbf{r}=\frac{\partial\Pi}{\partial t}k\mathbf{r} \tag{33a}$$

and Π is the solution of:

$$(\nabla^2+k^2)\Pi=-P_{0r}\mu/kr. \tag{33b}$$

IV. FORMAL SOLUTION

The solutions to (33b) valid inside and outside the earth are (with symmetry in φ)

$$\Pi\equiv\pi_2=\sum_{n=0}^{\infty}A_n j_n(k_2r)P_n(\cos\theta) \quad \text{(inside)} \tag{34a}$$

and

$$\Pi\equiv\pi_1=\sum_{n=0}^{\infty}R_n(\rho)P_n(\cos\theta) \quad \text{(outside)} \tag{34b}$$

$$(\rho\equiv k_0r=2\pi r/\lambda_0)$$

[15] Frank- v Mises, *Differentialgleichungen der Physik* (Mary S. Rosenberg, New York, 1943), vol. II, p. 872.

[16] H. Bremmer, Physica, **15**, 301 (1948).

where

$$\rho^2\frac{d^2R_n}{d\rho^2}+2\rho\frac{dR_n}{d\rho}+R_n\left\{\rho^2\left(\frac{k_1}{k_0}\right)^2-n(n+1)\right\}$$
$$=-\frac{\mu(n+\frac{1}{2})}{\rho k_1 k_0}\int_0^\pi P_r\rho^2 P_n(\cos\theta)\sin\theta d\theta. \quad (35)$$

For a doublet source located at $\theta=0$, $r=b$, designated ϱ_b,

$$P=P_0\delta(\varrho-\varrho_b)k_0^3 \quad (\varrho\equiv k_0\mathbf{r}) \quad (36)$$

where P_0 is the strength of the doublet, and δ is the Dirac "delta" function[17] which is defined in such a way that

$$\int_{\text{all space}}\delta(\varrho-\varrho_b)d\varrho=1,$$
$$\int_{\text{all space}}\delta(\varrho-\varrho_b)f(\varrho)d\varrho=f(\varrho_b). \quad (11a)$$

The inhomogeneous term must be retained in (35) since it may not be assumed here, as it was in the uniform atmosphere problem,[18] that the singularity in the Hertz potential at the source has the form e^{ik_1R}/R (see Fig. 2). The usual conditions of continuity of E_t and H_t, and the boundary condition at ∞ become

$$\text{(a)}\ k_1\pi_1=k_2\pi_2 \quad \text{at} \quad r=a \quad (\text{i.e. } \rho=k_0a=\rho_a),$$
$$\text{(b)}\ \frac{1}{k_1}\frac{\partial}{\partial r}(r\pi_1)=\frac{1}{k_2}\frac{\partial}{\partial r}(r\pi_2) \quad \text{at} \quad r=a, \quad (37)$$

(c) outgoing wave, with free-space velocity of light, for

$$r\longrightarrow\infty.$$

If one sets

$$\phi_n(s)=-\frac{P_0s^2\delta(\mathbf{s}-\varrho_b)}{sk_1}\int_0^\pi P_n(\cos\theta)\sin\theta d\theta \quad (38)$$

then the solution to (35) may be found in terms of a *Green's function* $G_n(\rho, s)$ by

$$R_n(\rho)=\int_0^\infty G_n(\rho, s)\phi_n(s)ds \quad (39)$$

where the Green's function for Eq. (35) is[19]

$$G_n(\rho, s)_{\rho\gtrless s}=\mp\frac{1}{2s^2}\left\{\frac{u_{1n}(s)u_{2n}(\rho)-u_{1n}(\rho)u_{2n}(s)}{u_{1n}(s)u_{2n}'(s)-u_{1n}'(s)u_{2n}(s)}\right\}$$
$$+B_nu_{1n}(\rho)+C_nu_{2n}(\rho). \quad (40)$$

[17] P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press, London, 1935), second edition, pp. 72–77.
[18] See reference 17, Van der Pol and Bremmer.
[19] E. L. Ince, *Ordinary Differential Equations* (Dover Publications, New York), Chapter XI, p. 257.

The u's are linearly independent solutions of the homogeneous form of (35) and the B_n and C_n are to be adjusted so that $G_n(\rho, s)$ satisfies the conditions (37).

A useful simplification is attained by use of Abel's identity[20] which states that if

$$\Delta\equiv u_1(x)u_2'(x)-u_1'(x)u_2(x), \quad (41)$$

where the u's are solutions of

$$p_0(x)u''(x)+p_1(x)u'(x)+p_2(x)u(x)=0 \quad (42)$$

then

$$\Delta=\Delta_0\exp\left(-\int_{x_0}^x\frac{p_1}{p_0}dx\right) \quad (43)$$

where Δ_0, is the value of (41) at $x=x_0$. For the Eq. (35a)

$$\Delta=\Delta_0\left(\frac{s_0}{s}\right)^2 \quad (44)$$

and it is significant that this result is independent of the form of k_1.

Application of the boundary conditions (37) to (34a) and (34b), using (40) for R_n, will yield solutions for A_n, B_n, C_n, provided one identifies the *traveling* and *standing* wave solutions of (35). Thus, if $u_{2n}(\rho)$ is to represent the outgoing wave solution and $u_{1n}(\rho)$ the standing wave, (37c) implies that we want

$$G(\rho)\longrightarrow\text{const.}\times u_2(\rho) \quad \text{for} \quad \rho\longrightarrow\infty,$$

and this immediately enables us to solve for B_n from (40), i.e.

$$B_n=\frac{1}{2\Delta_0s_0^2}u_{2n}(s). \quad (45)$$

If the corresponding solution for C_n is inserted in (40),

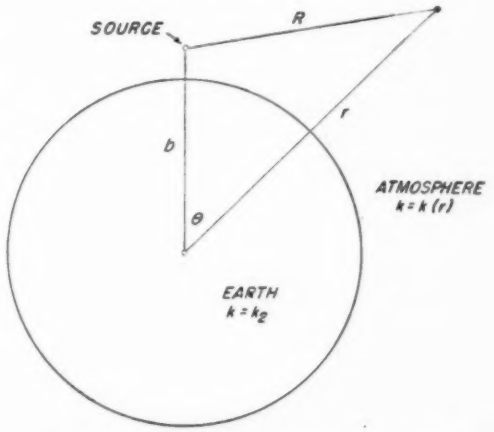


FIG. 2. Radiating doublet above the earth.

[20] See reference 19, Chapter V.

the Green's function is:

$$G_n(\rho, s)=u_{1n}(\rho)\frac{u_{2n}(s)}{\Delta_0 s_0^2}$$

$$-u_{2n}(\rho)\frac{u_{2n}(s)u_{1n}(\rho_a)}{\Delta_0 s_0^2\cdot u_{2n}(\rho_a)}\cdot M_n \quad \rho_a<\rho<s \quad (46)$$

$$=\frac{u_{2n}(\rho)}{\Delta_0 s_0^2}\left\{u_{1n}(s)-\frac{u_{2n}(s)u_{1n}(\rho_a)}{u_{2n}(\rho_a)}\cdot M_n\right\} \quad \rho>s$$

where

$$M_n\equiv\frac{\dfrac{j_n(k_2a)+k_2aj_n'(k_2a)}{j_n(k_2a)}-\dfrac{k_2^2}{k_1^2}\dfrac{u_{1n}(\rho_a)+k_0au_{1n}'(\rho_a)}{u_{1n}(\rho_a)}}{\dfrac{j_n(k_2a)+k_2aj_n'(k_2a)}{j_n(k_2a)}-\dfrac{k_2^2}{k_1^2}\dfrac{u_{2n}(\rho_a)+k_0au_{2n}'(\rho_a)}{u_{2n}(\rho_a)}}.$$

To illustrate the application of this result, consider $k_1=k_2=k_0$, corresponding to a dipole in free space. We then find

$$\begin{aligned}u_{1n}(\rho)&=j_n(\rho),\\ u_{2n}(\rho)&=h_n'(\rho),\\ \Delta_0 s_0^2&=i.\end{aligned} \quad (47)$$

Evaluation of (39), and subsequent substitution into (34b) and (33a) gives

$$\mathbf{A}=\sum_{n=0}^{\infty}\mathbf{r}\frac{2n+1}{4\pi i}\frac{\dot{P}_0\mu k_0^2}{\rho_b}j_n(\rho)h_n^{(1)}(\rho_b)P_n(\cos\theta)$$

$$\rho<\rho_b$$

$$=\sum_{n=0}^{\infty}\mathbf{r}\frac{2n+1}{4\pi i}\frac{\dot{P}_0\mu k_0^2}{\rho_b}j_n(\rho_b)h_n^{(1)}(\rho)P_n(\cos\theta) \quad (48)$$

$$\rho>\rho_b$$

$$=\frac{\dot{P}_0\mu}{4\pi b}\frac{e^{ik_0R}}{R}\mathbf{r}$$

where R is measured from the source. This is the spherical vector potential for a doublet. The usual expressions for the field vectors may be found upon application of (18) and (19).

For a dipole above an earth surrounded by free space, $k_1=k_0\neq k_2$, the u's are given by (47), and the vector potential $\mathbf{A}$ reduces to that given by Van der Pol and Bremmer.[18]

For a refracting atmosphere one may write, in general,

$$n^2=k_1^2/k_0^2=1+f(\rho). \quad (49)$$

A typical curve of index *vs.* height is Fig. 3. It must be borne in mind that $f(\rho)\to 0$ for $\rho\to\infty$.

If $f(\rho)$ is (mis-) represented by a function of the form $A'+B'/\rho^2$, then one may obtain exact solutions to (35)

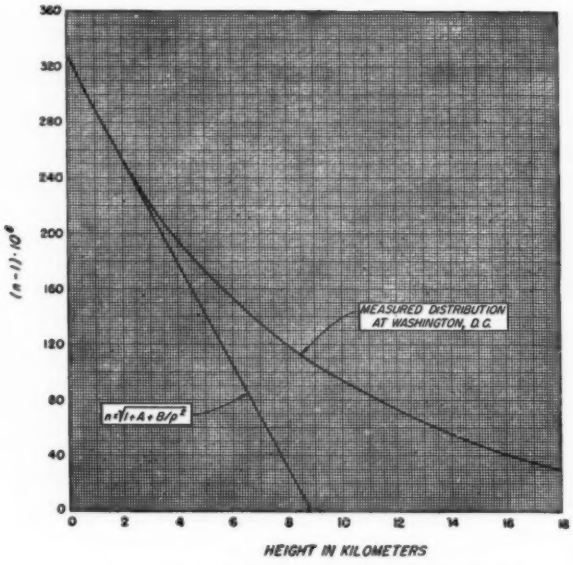


FIG. 3. Index of refraction *vs.* height above the earth.

in terms of Bessel functions, and such a solution has been considered by Ponomarev.[21] There arises the question of how closely the solution for this hypothetical atmosphere will correspond to the field strengths in the actual atmosphere. One might argue qualitatively, that since the index of refraction of this hypothetical atmosphere is such that it postulates index gradients that are too large (see Fig. 3), it will overestimate the amount of energy "reflected" back to the earth.

To treat a more general $f(\rho)$, we note that the most commonly used approximate solutions to equations of the form

$$u''(\rho)+\frac{2}{\rho}u'(\rho)+u\left[1+f(\rho)-\frac{n(n+1)}{\rho^2}\right]=0 \quad (50)$$

are found by the *WKB* method[6] which gives

$$u_{\pm}=\frac{A_{\pm}}{\rho\{1+f(\rho)-[n(n+1)/\rho^2]\}^{\frac{1}{4}}}$$

$$\times\exp\left\{\pm i\int_{\rho_1}^{\rho}\left[1+f(\rho)-\frac{n(n+1)}{\rho^2}\right]^{\frac{1}{2}}d\rho\right\}, \quad (51)$$

where ρ_1 represents the zero of the integrand. The use of the *WKB* solution here should not be confused with the phase-integral technique of Eckersley and Millington, wherein n is a parameter to be determined by the boundary conditions. We have already satisfied the boundary conditions, and the n's represent positive integers. There are, however, several objections to the use of *this WKB* solution for the problem under consideration. The solution (51) is not valid near the point ρ_1. Furthermore, the values of A_+ and A_- change as ρ

[21] M. I. Ponomarev, Bull. Acad. Sci. U.S.S.R., Technical Series, No. 9, p. 219 (1946).

circles the point ρ_1 in the complex plane, because the functions $u_\pm$ are not single valued, whereas (50) should have a single valued solution near the ordinary point ρ_1. The connection formulas relating the values $A_\pm$ in the various regions of the complex plane are discussed by Furry.[22] It has also been noted by Langer[23] in his treatment of the radial equation involved in the application of the Schrödinger wave equation to central field problems, that the *WKB* technique may not be applied directly to an equation of the form (50).

These difficulties were all overcome by Langer.[23] Upon applying his techniques to Eq. (50) the solutions are found to be:

$$u_{\frac{1}{2}n}(\rho)\cong\frac{1}{\rho_1^{\frac{1}{2}}\{1+f(\rho)-[(n+\frac{1}{2})^2/\rho^2]\}^{\frac{1}{4}}}\times\left\{\int_{\rho_1}^{\rho}\{1+f(\rho)-[(n+\frac{1}{2})^2/\rho^2]\}^{\frac{1}{2}}d\rho\right\}^{\frac{1}{2}}\times\frac{J_{\frac{1}{3}}}{H_{\frac{1}{3}}}\left\{\int_{\rho_1}^{\rho}\{1+f(\rho)-[(n+\frac{1}{2})^2/\rho^2]\}^{\frac{1}{2}}d\rho\right\}. \quad (52)$$

These are valid in all regions in which the *WKB* solutions hold, in addition to the region near and including the zero point of the radical, and are single valued. One might note the appearance of the $(n+\frac{1}{2})^2$ to replace the $n(n+1)$ which appears in the usual *WKB* solution and is known to be in error. The functions (52) are not good approximations at zeros of the radical other than ρ_1.

If $f(\rho)$ is set $=0$ in (50) the solutions should represent Bessel functions, and indeed, the expressions (52) reduce to the "Watson" or "Hankel" approximation to the Bessel function,[24] usually obtained by a saddle-point approximation to the integral representation of the Bessel function.

In the next section, the integrals involved in (52) will be carried out for the standard atmosphere of Fig. 3.

V. FIELD STRENGTH CALCULATION—STANDARD ATMOSPHERE

The field will be calculated at the earth ($\rho=\rho_a$) and the earth will be considered to be a good conductor ($k_2\to\infty$), in which case (34b) becomes

$$\pi_1=\sum_{n=0}^{\infty}-\frac{(2n+1)P_0\mu k_0^2}{4\pi\rho_b k_1(b)}\cdot\frac{u_{2n}(\rho_b)}{u_{2n}'(\rho_a)}P_n(\cos\theta), \quad (53)$$

where the u's are given by (52).

Having obtained the series (53) one may directly apply the Watson[2] summation technique whereby π_1 is expressed as

$$\pi_1=-\frac{P_0\mu k_0^2}{4\pi\rho_b k_1(b)}\oint\frac{u_{2\nu-\frac{1}{2}}(\rho_b)}{u_{2\nu-\frac{1}{2}}'(\rho_a)}\frac{P_{\nu-\frac{1}{2}}(-\cos\theta)}{i\cos\nu\pi}d\nu \quad (54)$$

where the contour consists of small circles around the half-odd integers on the positive real axis. This contour is then shown to be equivalent to one encircling the first quadrant, so that one may set

$$\pi_1=-\frac{P_0\mu k_0^2}{4\pi\rho_b k_1(b)}\sum_s\frac{u_{2\nu_s-\frac{1}{2}}(\rho_b)}{\cos\nu_s\pi}\times\frac{P_{\nu_s-\frac{1}{2}}(-\cos\theta)}{[(\partial/\partial\nu)u_{2\nu-\frac{1}{2}}'(\rho_a)]_{\nu=\nu_s}}, \quad (55)$$

where the ν_s are determined by

$$Rl(\nu)>0,\qquad u_{2\nu_s-\frac{1}{2}}'(\rho_a)=0,\qquad Im(\nu)>0. \quad (56)$$

The justification for the deformations of the contour in (54) which lead to (55) follows almost exactly as in the homogeneous atmosphere case.[2]

Anticipating that the low order (i.e., smallest ν) roots of (56) will correspond to small values of the argument of the Hankel function of (52), and will involve integrals over small heights above the earth, we restrict ourselves to the region in which $f(\rho)$ is linear, i.e., $[(\rho_1-\rho_a)/k_0]$ <3 k.m. (See Fig. 3). We find

$$u_{2\nu_s-\frac{1}{2}}(\rho)\cong\frac{e^{i\pi/2}}{\rho}\sqrt{\tfrac{2}{3}}\delta^{\frac{1}{2}}H_{\frac{1}{3}}^{(1)}\left[\frac{(1+A)^{\frac{1}{2}}}{(\rho)^{\frac{1}{2}}}e^{3i\pi/2}\frac{2^{\frac{1}{2}}}{3}\delta^{\frac{3}{2}}\right], \quad (57)$$

where

$$A\equiv\frac{\rho_a}{2}f'(\rho_a)+f(\rho_a).$$

$$\delta\equiv\left(\frac{\nu^2-B}{1+A}\right)^{\frac{1}{2}}-\rho\ \left(\frac{\delta}{\rho}\ll 1\right). \quad (58)$$

$$B\equiv-\frac{\rho_a^3}{2}f'(\rho_a)$$

The roots of (56) are then

$$\left(\frac{\nu_s^2-B}{1+A}\right)^{\frac{1}{2}}=\rho_a+\rho_a^{\frac{1}{3}}\tau_s(1+A)^{-\frac{1}{3}} \quad (59)$$

where

$$\tau_0=0.808e^{i\pi/3},$$

$$\tau_1=2.58e^{i\pi/3}$$

and an extensive table of τ_s is given in Van der Pol and

[22] W. H. Furry, Phys. Rev. **71**, 360–371 (1947).
[23] R. E. Langer, Phys. Rev. **51**, 628 (1937).
[24] See, for example, W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949).

Bremmer.[18] For ν_s in the first quadrant we also use

$$P_{\nu_s-\frac{1}{2}}(-\cos\theta)\sim\frac{1}{(2\pi\nu_s\sin\theta)^{\frac{1}{2}}}\times\exp[-i\nu_s(\pi-\theta)+i\pi/4], \quad (60)$$

$$\frac{1}{\cos\nu_s\pi}\sim 2e^{i\nu_s\pi}.$$

Inserting (58–60) into (55) and lowering the transmitter to the earth ($\rho_b=\rho_a$) leads to

$$\pi_1\cong\sum_s\frac{P_0\mu k_0^2}{4\pi k_1(b)}\frac{\exp[i\pi/4+2i\nu_s\pi+iRl(\nu_s)\theta]}{\rho_a^{\frac{1}{2}}(2\pi\sin\theta)^{\frac{1}{2}}\tau_s}\times\left\{\frac{(1+A)^{\frac{1}{4}}}{(1+A+B/\rho_a^2)^{\frac{1}{4}}}\right\}\times\exp\left[-(1+A)^{\frac{1}{4}}\frac{\sqrt{3}}{2}|\tau_s|\frac{\rho_a^{\frac{1}{2}}\theta}{(1+A+B/\rho_a^2)^{\frac{1}{4}}}\right]. \quad (61)$$

A comparison of this result with that for a homogeneous free-space atmosphere is readily achieved by setting $A=B=0$. The distance dependence is contained in the last factor in each term. Clearly, the major effect of atmospheric refraction is to alter the rate of expotential decay of field strength with distance, which for free space takes the form

$$\exp(-\sqrt{\tfrac{3}{2}}|\tau_s|D/\rho_a^{\frac{1}{2}}) \quad (62)$$

where D refers to distance along the earth. For large D, the $s=0$ term predominates, and the behavior of $f(\rho)$ at large heights has practically no influence on the field strength since it only enters into the calculation of the τ_s with large s. This substantiates the intuitive notion that the energy is "fed" into the diffraction region by those "rays" which are almost horizontal. (Again, these arguments hold only for monotonic $f(\rho)$, in which case the τ_s will increase steadily.)

To get (61) into the free-space form, we introduce an "effective earth's radius" ρ_{eff}, and a new θ coordinate θ', and require

$$\rho_{\text{eff}}\theta'=\rho_a\theta=D. \quad (63)$$

It follows then that if we set

$$\rho_{\text{eff}}=\frac{\rho_a}{1+(A/4)-(3B/4\rho_a^2)}=\frac{\rho_a}{[1+(n-1/2)+\rho(dn/d\rho)]_{\rho=\rho_a}} \quad (64)$$

the Hertz potential, Eq. 61, corresponds to propagation through free space over an enlarged earth, thus verifying the validity of the concept.

Reference to Eq. (16) indicates that the value predicted for this effective radius by the ray theory is not reliable. For example, if the index of refraction n behaves as

$$n=1+\alpha+\beta(\rho-\rho_a) \quad (65)$$

then (64) gives

$$\rho_{\text{eff}}\cong\frac{\rho_a}{1+\alpha/2+\rho_a\beta}\ \text{(wave theory)},$$

whereas

$$\rho_{\text{eff}}\cong\frac{\rho_a}{1+(\rho_a\beta/1+\alpha)}\ \text{(ray theory)}. \quad (66)$$

For the standard atmosphere, however, $A\cong-B/\rho_a^2$, so that (64) will give the same value for the effective earth's radius (for large D) as predicted by the ray theory for almost horizontal rays.

Although the first mode, corresponding to $s=0$ in (61), accurately represents the field at large distances, it is necessary to sum many such modes as one gets nearer the transmitter. To extend the effective radius concept to regions closer to the transmitter, one need only note that the arguments outlined here may be applied termwise to each of the modes. This follows from the fact that the phase factor of each term, may be treated in exactly the same way as was the attenuation factor, so that the relative phases of the various modes are maintained. The number of modes for which the representation (57) is valid will be determined by the condition

$$(\rho_1-\rho_a)/k_0=r_1-a\leq 3\ \text{k.m.} \quad (67)$$

as may be seen from Fig. 3. We find from (56)–(59) that

$$r_{1_s}=r_a+(\lambda_0/2\pi)^{\frac{1}{3}}106x_s^{\frac{1}{3}} \quad (68)$$

where x_s is the sth root of

$$J_{2/3}(x)=J_{-2/3}(x). \quad (69)$$

For $\lambda_0=1$ meter, there will be of the order of a hundred modes within the 3 k.m., and correspondingly more for shorter wave-lengths.

Conclusion

There is ample justification for the use of an enlarged earth-homogeneous atmosphere model for calculating the fields existing under standard refraction conditions. Any measured deviation of field strength from that calculated on this basis is fairly conclusive evidence that the atmosphere has departed from standard conditions.

Acknowledgment

I would like to express my thanks to Dr. T. J. Carroll for his extremely helpful advice and encouragement, and Mr. George A. Hufford for his careful reading and criticism of the manuscript.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Experimental Measurement of Incompressible Flow Along a Cylinder with a Conical Nose

E. V. LAITONE
University of California, Berkeley, California
(Received July 7, 1950)

The experimental pressure distribution data from low speed wind tunnel tests at zero angle of attack on two different conical noses attached to a cylindrical after-body are compared with theory. The simplified theoretical equation used is a higher order approximation valid up to the order $r^4 \ln r$. It is shown that the conical nose with a vertex angle of 20 degrees is in agreement but the larger cone with a vertex angle of 30 degrees shows a discrepancy between the measured pressures and the approximate theory. In both cases, however, the theory is in excellent agreement with the surface pressures on the cylindrical after-body.

EXPERIMENTAL PROCEDURE

THE wind tunnel tests were performed in the 3×3 foot open throat wind tunnel at the University of California at Berkeley by Messrs. N. N. Cahan, R. B. Glickman, G. B. Himel, and J. L. Taylor.

The models were made of aluminum and consisted of a conical nose in front of a 12-in. long, 4-in. diameter, cylindrical after-body as shown in Fig. 1. Two conical noses were tested; one with a 20-degree vertex angle and the other with a 30-degree vertex angle. The models were supported at zero angle of attack along the tunnel center line by means of a 2-in. diameter sting faired into a vertical support strut as indicated in Fig. 1. The tests were all made at a wind velocity of 85 miles per hour corresponding to a Mach number of 0.10 and a Reynolds number of 250,000 based on the 4-in. diameter of the cylindrical after-body.

The surface pressures were measured by means of $\frac{1}{16}$-in. diameter orifices connected to an alcohol multiple tube manometer and are shown in coefficient form in Figs. 2 and 3. Each test point shown is the average of 4 simultaneous readings taken at 90-degree intervals on the circumference of each longitudinal station. This effectively corrected the surface pressures to axial flow at zero angle of attack since the variation in the circumferential readings of the pressure indicated the local angle of attack was never greater than 0.1 degree from zero. The surface pressures were also corrected for wind tunnel boundary and support interference.

SYMBOLS

a = free stream velocity of sound.
U = free stream velocity.
M = free stream Mach number $= U/a$.
$\beta = (1-M^2)^{\frac{1}{2}}$.
C_p = local surface pressure coefficient

$$= \frac{\text{pressure difference between surface and free stream}}{\text{free stream dynamic pressure}}$$

$$= \frac{\Delta p}{\frac{1}{2}\rho U^2} \approx -2\frac{\Delta u}{U} - (r_0')^2 \text{ (see reference 1).}$$

Δu = difference between local surface velocity and free stream velocity.
(x, r) = cylindrical coordinates (see Fig. 1).
r_0 = coordinate to surface of symmetrical body of revolution.
$r_0' = dr_0/dx$ = local surface slope.
θ = arc $\tan r_0' \approx r_0'$, semi-vertex angle of conical nose.
ξ = variable of integration along x axis when (x, r_0) is considered as a fixed local point where the velocity and pressure are being computed.
L = length of conical nose.
$S(x) = \pi r_0^2$ = cross-sectional area of body of revolution.
$S'(x) = dS/dx = 2\pi r_0 r_0'$.

The notation $R = O(r^n)$ means that the remainder term in the series expansion is such that $\lim_{r\to 0}(R/r^n)$ is bounded or zero. Also for a conical nose $O(r) = O(dr/dx) = O(\theta)$.

COMPARISON WITH THEORY

In reference 1, it is shown that the surface pressure coefficient on any body of revolution in subsonic flow is given by

$$C_p[x, r_0(x)] = \frac{1}{2\pi}\int_0^L \frac{(\xi - x)S'(\xi)\,d\xi}{[(x-\xi)^2 + (\beta r_0)^2]^{\frac{1}{2}}} - (r_0')^2 + O(r^4 \ln r). \quad (1)$$

The remainder term in Eq. (1) is exact only in the limit as $\beta = (1-M^2)^{\frac{1}{2}}$ approaches unity; that is, for incompressible flow.

Applying Eq. (1) to the models given in Fig. 1 we have from $x=0$ to $x=L$

$$S(x) = \pi x^2(\tan^2\theta) = \pi\theta^2 x^2 + O(\theta^4)$$
$$S'(\xi) = 2\pi\theta^2\xi + O(\theta^4)$$

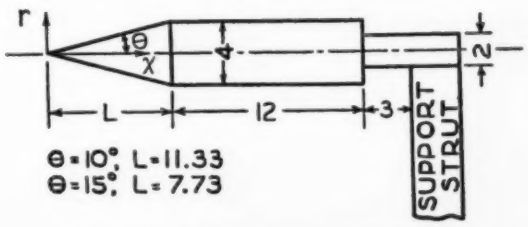


FIG. 1. Model dimensions and coordinates.

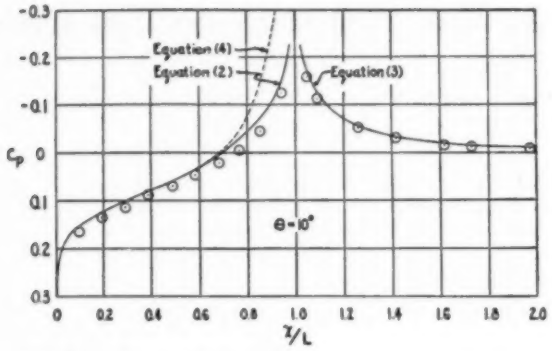


FIG. 2. Comparison of the measured surface pressure coefficient (⊙) with theory for a 20° conical nose.

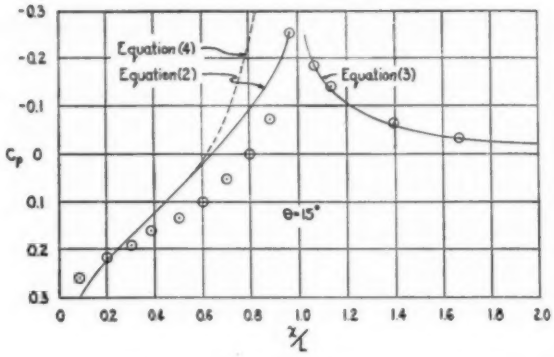


FIG. 3. Comparison of the measured surface pressure coefficient (⊙) with theory for a 30° conical nose.

and from $x=L$ to $x=\infty$

$$S'(\xi)=0$$

so that on the conical nose $(0<x<L)$

$$C_p=\theta^2\int_0^L\frac{(\xi-x)\xi d\xi}{[(\xi-x)^2+(\beta\theta x)^2]^{\frac{3}{2}}}-\theta^2+O(\theta^4\ln\theta)$$

$$=-\theta^2\left[\frac{1}{([1-(x/L)]^2+[\beta\theta(x/L)]^2)^{\frac{1}{2}}}\right.$$

$$\left.-\sinh^{-1}\left(\frac{1-(x/L)}{\beta\theta(x/L)}\right)-\sinh^{-1}\left(\frac{1}{\beta\theta}\right)+1\right] \quad (2)$$

and on the cylindrical after-body $(x>L)$

$$C_p=\theta^2\int_0^L\frac{(\xi-x)\xi d\xi}{[(\xi-x)^2+(\beta\theta L)^2]^{\frac{3}{2}}}+O(\theta^4\ln\theta)$$

$$=-\theta^2\left[\frac{1}{([(x/L)-1]^2+(\beta\theta)^2)^{\frac{1}{2}}}\right.$$

$$\left.+\sinh^{-1}\left(\frac{(x/L)-1}{\beta\theta}\right)-\sinh^{-1}\left(\frac{x/L}{\beta\theta}\right)\right]. \quad (3)$$

The theoretical variation of the surface pressure coefficient as given by Eqs. (2) and (3) is shown by the solid line in Figs. 2 and 3. It is apparent that the theory of reference 1 is satisfactory for the cone with the 20-degree vertex angle but not for the larger one with the 30-degree vertex angle. This indicates the limitations imposed by the approximation obtaining Eq. (1) in reference 1 by neglecting terms of the order $r^4\ln r$. It is seen, however, that the surface pressures on the cylindrical after-body are in excellent agreement in both cases.

If the terms containing θ in Eq. (2) are expanded to the order $\theta^3=0$ we obtain for the conical nose $(0<x<L)$

$$C_p=-\theta^2\left\{2\ln\frac{\beta\theta(x/L)}{2(x/L[1-(x/L)])^{\frac{1}{2}}}\right.$$

$$\left.+\frac{2-(x/L)}{1-(x/L)}\right\}+O(\theta^3). \quad (4)$$

This expression could also be obtained directly from Eq. 10 of reference 1. It is seen in Figs. 2 and 3 that the values given by Eq. (4) indicate a greater velocity increment than do those given by Eq. (2) which are already slightly greater than the experimental values.

A comparison of Figs. 2 and 3 with a distribution of the type given by $(x/L)^m$ shows that a velocity distribution varying with some power of (x/L) would provide a reasonable approximation only for the mid-portion of a cone with a very large vertex angle. This indicates that the boundary layer solutions given by W. Mangler and others (see reference 2) are not valid for ordinary slender cones as used in supersonic aircraft. However, for extremely slender cones, a constant velocity distribution may be assumed over the cone since Eq. (4) then approaches the limiting value

$$C_p=-2\theta^2\ln\theta+O(\theta^2), \quad (5)$$

which is valid for either subsonic or supersonic flow.

[1] E. V. Laitone, Quart. Appl. Math. 5, No. 2, 227–231 (July, 1947): (See also J. Aeronaut. Sci. 14, No. 11, 639 (November, 1947).

[2] W. Mangler, "Boundary layers on bodies of revolution in symmetrical flow," M.A.P. Volkenrode Interrogation Report Ref. MAP-VG46 (April 15, 1946).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Graphical Determination of the Fermi Level in a Simple Impurity Semiconductor

K. LEHOVEC AND H. KEDESDY
Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey
(Received July 12, 1950)

A graphical determination of the position of the Fermi level in a simple impurity semiconductor at zero space charge is discussed. Measurements by G. Busch and H. Labhard indicate that the effective masses of electrons and holes in SiC crystals are of the order of mass of a free electron.

I. INTRODUCTION

THE electrons which carry the current in an n-type impurity semiconductor are supplied by thermal ionization of impurities (called donors). In the simplest case there is only one kind of impurity. For the calculation of the electrical properties one needs to know (1) the energy states which can be occupied by electrons and (2) the distribution of electrons in these energy states. The distribution of electrons in thermal equilibrium is given by the Fermi distribution function in which the Fermi energy (electrochemical potential, free energy per added electron) is the characteristic parameter.

An equation for the determination of the Fermi level in thermal equilibrium and at zero space charge for the simple model of an impurity semiconductor mentioned above, has been given by K. Shifrin.[1] This equation has been applied by E. H. Putley[2] to an explanation of measurements of electrical conductivity of germanium. In both papers there is missing a graphical representation of results enabling a quick determination of the Fermi level as a function of the concentration of impurities, the ionization energy of impurities, and temperature. Since such a determination is a task frequently encountered in dealing with semiconductors, it was considered worth while to publish graphs permitting a convenient determination of the Fermi level. The same graphs may be used for p-type impurity semiconductors with one kind of impurity. For reasons of clarity the derivation of the position of the Fermi level is briefly reviewed.

The spectrum of energy states under discussion (see Fig. 1) consists of the energy band of the basic lattice, to which electrons are released ("conduction band") and of the energy level of the impurities, from which the electrons are released by thermal agitation. The difference between the donor energy level and the conduction band is the ionization energy E. The number of energy states per unit volume of the impurity level is equal to the concentration of impurities N. The number of energy states[3] in the conduction band with energies between ϵ and $\epsilon+d\epsilon$ is

$$V\cdot D(\epsilon)d\epsilon=4\pi(2m^*/h^2)^{\frac{3}{2}}\epsilon^{\frac{1}{2}}\cdot V\cdot d\epsilon \qquad (1)$$

(ϵ=energy of an energy state relative to the band boundary as shown in Fig. 1; h=Planck's constant; V=volume of the crystal).

Equation (1) holds generally for energy values near a band boundary. The peculiarities of the crystal lattice under consideration express themselves by the effective mass m^*, which may differ from the mass of a free electron.

In the case of thermal equilibrium the probability for occupation of an energy state[4] ϵ by an electron of given spin is obtained from the Fermi distribution function

$$f(\epsilon)=\frac{1}{\exp[(\epsilon-\xi)/kT]+1}, \qquad (2)$$

where ξ is the so-called Fermi energy, k=Boltzmann constant, and T=absolute temperature. Accordingly $1-f(\epsilon)$ is the probability for an energy state ϵ to be *un*occupied by an electron in thermal equilibrium.

II. DETERMINATION OF THE FERMI LEVEL

The position of the Fermi level may be expressed from considerations of the space charge. In the case of the simple energy scheme shown in Fig. 1, there is macroscopically zero space charge if concentration of unoccupied impurity levels equals the concentration of electrons in the conduction band.

This may be expressed by the equation

$$N[1-f(E)]=\int_{\epsilon=0}^{\infty} f(\epsilon)\cdot D(\epsilon)d\epsilon. \qquad (3)$$

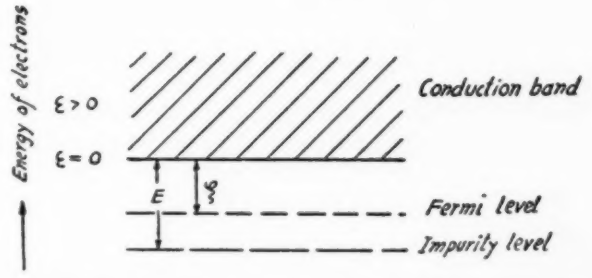
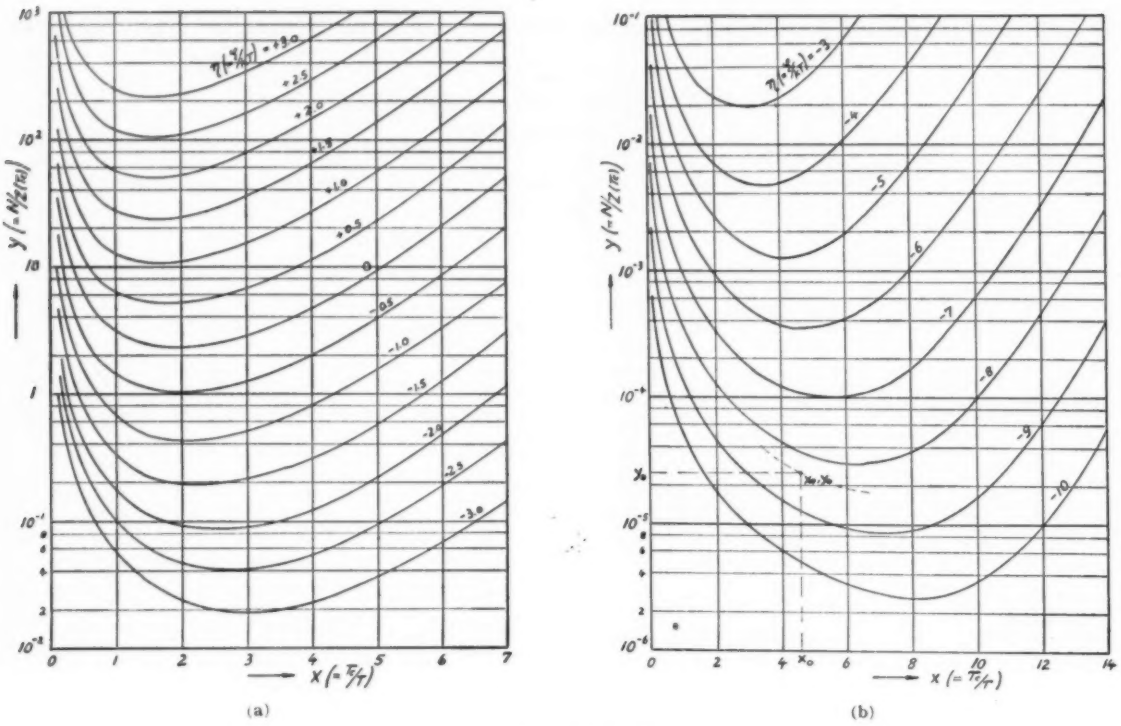


FIG. 1. Simple energy scheme of an impurity semiconductor.

[1] K. Shifrin, J. Phys. U.S.S.R. **8**, 242 (1944).
[2] E. H. Putley, Proc. Phys. Soc. (London) **A62**, 284 (1949).
[3] See, for instance, F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), p. 143 ff. A factor 2, which results from the two spin probabilities of an electron is included in Eq. (1).
[4] For electrons at the impurity level, ϵ has to be replaced by E.



(a) (b)

FIG. 2a, b. The function
$y(x)=(e^{x}\cdot e^{\eta}+1)\cdot 2\pi^{-\frac{1}{2}}\cdot F_{\frac{1}{2}}(\eta)\cdot x^{-\frac{1}{2}}$,
where
$$F_{\frac{1}{2}}(\eta)=\int_0^\infty \frac{t^{\frac{1}{2}}}{e^{t-\eta}+1}dt$$
(explanation of symbols in text).

The replacement of the upper limit in the integral by infinity does not introduce any appreciable error because of the smallness of the Fermi distribution function for $(\epsilon-\xi)/kT\gg 1$.

Equation (3) may be transformed with use of Eq. (1) and Eq. (2) to

$$N/Z(T)=[\exp(|E|/kT)\cdot\exp(\xi/kT)+1]\cdot 2\pi^{-\frac{1}{2}}\cdot F_{\frac{1}{2}}(\xi/kT), \quad (4)$$

where

$$F_{\frac{1}{2}}(\eta)=\int_0^\infty \frac{t^{\frac{1}{2}}}{\exp(t-\eta)+1}dt \quad (5)$$

(η substituted for ξ/kT and t substituted for ϵ/kT) and

$$Z(T)=2(2\pi m^{*}kT/h^{2})^{\frac{3}{2}}. \quad (6)$$

Numerical values of the function (5) have been tabulated by J. Mc Dougall and E. C. Stoner.[5]

Equation (4) connects four variables N, T, $|E|$, and ξ. The number of variables may be reduced to three by introduction of the new dimensionless variables

$$y=N/Z(T_c);\quad x=T_c/T;\quad \eta=\xi/kT, \quad (7a, b, c)$$

where $T_c=|E|/k$. Then Eq. (4) simplifies to

$$y=(\exp x\cdot\exp\eta+1)\cdot 2\pi^{-\frac{1}{2}}\cdot F_{\frac{1}{2}}(\eta)\cdot x^{-\frac{1}{2}}, \quad (8)$$

which can be represented by a family of curves.

III. GRAPHICAL REPRESENTATION AND DISCUSSION

In Fig. 2a, b, y is plotted as a function of x with η as a parameter. One is interested mostly in the range $x=T_c/T=|E|/kT\gtrsim 1$ (but not $\gg 1$) of the abscissa values, where the ionization energy E of the impurities is larger (but not too much larger) than the thermal energy kT.

The position of the Fermi level as a function of temperature T, concentration N, and ionization energy E of impurities may be found in the following manner. Consider for example the point $x_0(=|E|/kT)$, $y_0(=N/Z(T_c))$ (Fig. 2b) and determine the η-value η_0 of the curve passing through this point, if necessary by linear interpolation. From this η_0-value the Fermi level $\xi_0(=\eta_0\cdot kT)$ may then be computed.

A positive or negative value of η indicates that the Fermi level lies in or below the conduction band, respectively. If $\eta=0$, the Fermi level lies at the boundary of the conduction band.

In order to show the change of the Fermi level with temperature in Fig. 3, η is used as ordinate, x as ab-

[5] Phil. Trans. Roy. Soc. (London) **237**, 67 (1938).

scissa, and y as parameter. It is seen that η passes through a maximum. The maximum values of η lie on the dotted curve, which may be derived from Eq. (8) by partial derivation (since the maximum of η is to be determined: $d\eta/dx=0$; and since y is considered a constant parameter in Fig. 3, $dy/dx=0$) with respect to x. This gives[6]

$$\eta=-[x+\ln(\tfrac{2}{3}x-1)]. \tag{9}$$

The electron gas in the conduction band is called nondegenerate if the probability $f(\epsilon)$ for occupation of any of its energy states is very small in comparison with unity. This is the case if $\exp[(\epsilon-\xi)/kT]\gg 1$ (see Eq. (2)). Since $\epsilon\gtrsim 0$ for energy states in the conduction band, we shall define[7] nondegeneracy by $\eta=\xi/kT\lesssim 0$. It is perhaps worth while mentioning that in the case of nondegeneracy the occupation probability of energy states in the conduction band reduces to the Boltzmann factor $\exp[-(\epsilon-\xi)/kT]$ and that the integral $2\pi^{-\frac{1}{2}}\cdot F_{\frac{1}{2}}(\eta)$ becomes e^{η}. The temperature range where the electron gas is degenerate or accordingly $\eta>0$ may be seen readily from Fig. 3. Maximum degeneracy (if any) occurs at the temperature corresponding to the abscissa value of the maximum of the η-curve under consideration. For $y\leq 2.3$ the electron gas is nondegenerate at all temperatures; introduction of numerical values shows that in this case

$$N/|E|^{\frac{3}{2}}\leq 1.4\times 10^{22}\cdot(m^*/m)^{\frac{3}{2}}, \tag{10}$$

where N is expressed in cm^{-3} and E in electron volts, and m^*/m is the effective mass expressed in units of free electron mass.

IV. APPLICATION OF THE GRAPH $\eta=f(x, y)$ FOR DETERMINATION OF THE EFFECTIVE ELECTRON MASS m^* IN SILICON CARBIDE

It is common practice to determine E and NZ (or accordingly $Nm^{*\frac{3}{2}}$) from measurements of the Hall constant as function of temperature for a semiconductor of the simple scheme of Fig. 1. For separate determination of N and m^* still another combination (different from $Nm^{*\frac{3}{2}}$) must be known. Such a combination $(N\cdot E^{-\frac{1}{2}}\cdot m^{*-\frac{1}{2}})$ may be obtained from $y=N/Z(T_c)$ (see Eq. (6) and recall that $T_c=|E|/k$). The parameter y can be determined from the x-value corresponding to maximum η, since the latter must lie on the dotted curve of Fig. 3. Considering that x is determined by temperature T and ionisation energy E (Eq. 7), it remains to deduce the temperature at which η has its maximum from measurements.

For this purpose measurements of conductivity as function of temperature can be used, if the electron mobility depends[8] on temperature as $T^{-\frac{1}{2}}$. It can be shown[9] that in this case a maximum of conductivity occurs at the same temperature at which η has its maximum.

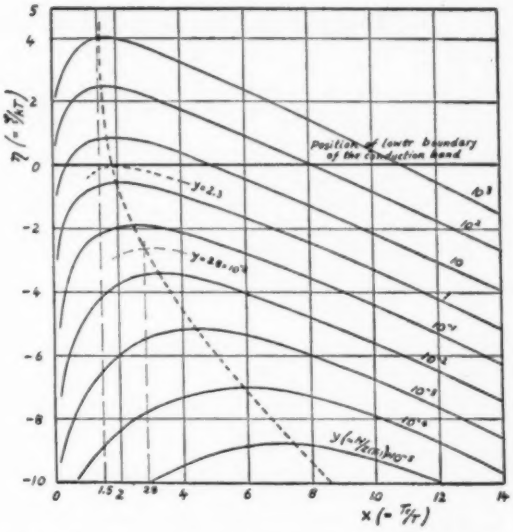


FIG. 3. The function $\eta(x)$ with y as parameter. The dotted curve $\eta=-[x+\ln(\frac{2}{3}x-1)]$ connects the maximum values of the η-curves for various values of y. A positive η-value indicates that the Fermi level lies in the conduction band and accordingly, that the electron gas is degenerated.

The considerations above seem applicable to measurements by G. Busch and H. Labhard[10] on SiC crystals at higher temperatures.[11] The product of Hall constant and conductivity (which should be proportional to the mobility if there is electron conduction only) varies with temperature as $T^{-\frac{1}{2}}$. From the temperature dependence of the Hall constant (see Fig. 5 of G. Busch and H. Labhard) one finds for crystal No. 7 for instance, that $E=0.24$ ev and NZ (400°K) $=1.3\times 10^{40}$ cm^{-6}. Maximum conductivity occurs at $T_0=1000$°K. Hence, $x=|E|/kT_0=2.8$ and $y=2.8\times 10^{-2}$ (see Fig. 3). From the values of E, NZ, and y the concentration of impurities $N=8\times 10^{19}$ cm^{-3} and the effective mass $m^*/m=2.5$ is obtained by using Eqs. (6) and (7a). An analogous calculation for other crystals (both p- and n-type) measured by G. Busch and H. Labhard yields similar values of m^*/m.

[6] Equation (9) is identical with Eq. (7) of K. Shifrin (reference 1) and Eq. (10) of E. H. Putley (reference 2).

[7] It should be kept in mind, however, that the transition from degeneracy to nondegeneracy is a gradual one.

[8] This temperature dependence of mobility is rather common for semiconductors and results from a mean free path which is independent of the electron energy. This is the case for scattering of electrons by lattice vibrations of the acoustical branch.

[9] Conductivity depends on the product of the concentration of electrons and their mobility. The concentration of electrons in the conduction band is given by the right side of Eq. (3), which can be transformed with the aid of Eq. (1) into $Z(T)\cdot 2\pi^{-\frac{1}{2}}\cdot F_{\frac{1}{2}}(\eta)$. The product $Z(T)\times$mobility becomes independent of temperature, if the mobility is proportional to $T^{-\frac{1}{2}}$. Hence the conductivity depends on the temperature only through the parameter η.

[10] G. Busch and H. Labhard, Helv. Phys. Acta **19**, 464 (1946).

[11] The simple energy scheme of Fig. 1 is not sufficient to explain the temperature dependence of conductivity and Hall constant at low temperature.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Expandability of a Wave-Guide Field in Terms of Normal Modes

J. VAN BLADEL*
The University of Wisconsin, Madison, Wisconsin
(Received July 17, 1950)

This article justifies the fact that the fields in a wave guide can be expanded as a linear combination of normal modes. The proof consists in showing that (a) this expansion is possible for the z components of the fields and modes and (b) that, in three dimensions, the fields as a whole are equal to a linear expansion of modes with precisely the same coefficients as were found in the expansion of the z components.

THE purpose of the present note is to give rigorous proof for the expansion of a wave-guide electromagnetic field in terms of the normal modes of the guide. It is assumed that the guide is filled with a homogeneous perfect dielectric, has perfectly conducting walls, and that the phenomena are harmonic in time.

It is necessary for the purpose of this proof to summarize the main properties of the normal modes. Such modes[1] are obtained by assuming a z dependence of the form e^{jhz} for the six components of the field vectors $\bar{E}, \bar{H}$ (see Fig. 1). The xy dependent part of a component will be indicated by brackets so that, for example,

$$E_z(x, y, z, t) = e^{j(hz+\omega t)}[E_z(x, y)]. \tag{1}$$

$\bar{E}$ and $\bar{H}$, on the other hand, satisfy Maxwell's equations

$$\text{curl}\bar{E} = -j\omega\mu\bar{H}, \tag{2}$$

$$\text{curl}\bar{H} = \bar{J} + j\omega\epsilon\bar{E}, \tag{3}$$

$$\text{div}\bar{H} = 0, \tag{4}$$

$$\text{div}\bar{E} = \rho/\epsilon, \tag{5}$$

written for free space conditions, i.e., $\bar{J} = \rho = 0$. They must also satisfy the boundary conditions at the wall. As a result of these conditions, only certain values of h are permissible, each of which defines a mode belonging to one of the following three classes:

(a) E (or TM) modes, which have zero H_z, and of which $[E_z]$ satisfies

$$\nabla^2[E_{zn}] + k_n^2[E_{zn}] = 0, \tag{6}$$

and $[E_{zn}]$ vanishes at the wall C. This allows an infinite, discrete set of values for k_n^2, whereby corresponding values of h^2 are given by

$$h_n^2 = \omega^2/c^2 - k_n^2. \tag{7}$$

The E mode, as a whole, is comprised of the vectors $\bar{E}_n^E$ and $\bar{H}_n^E$, the components of which can be expressed as a function of $[E_{zn}]$. The important property for the present purpose is, however, that the set of functions $[E_{zn}]$ is closed and complete.[2] This means that any reasonably continuous function, such as the components of an electromagnetic field, can be expanded in terms of the $[E_{zn}]$.

(b) H (or TE) modes, which have zero E_z, and of which $[H_z]$ satisfies (6) and $\partial[H_z]/\partial n$ vanishes at the wall C. This defines a discrete set of values for h_m^2 and $[H_{zm}]$, where the latter functions also form a closed and complete set.[2] The H mode is comprised of the vectors $\bar{E}_m^H$ and $\bar{H}_m^H$.

(c) TEM modes, which have zero E_z and H_z, and exist only when the dielectric region inside the guide is multiply connected. The propagation constant h is then equal to ω/c, and $[\bar{E}]$ is perpendicular to the wall and satisfies

$$\nabla_{xy}{}^2[\bar{E}] = 0. \tag{8}$$

In order to investigate the expandability, consider the field $\bar{E}, \bar{H}$ existing between the two cross sections S_1 and S_2 (Fig. 2), which enclose a certain current distribution $\bar{J}(x, y, z)$. Let $\bar{E}_p, \bar{H}_p$ be the field which would exist in a cavity obtained by metallizing S_1 and S_2. This field depends only upon $\bar{J}$ and vanishes in the absence of sources between S_1 and S_2. E_z and E_{zp} satisfy the equation

$$\nabla^2 E_z + \frac{\omega^2}{c^2}E_z = -j\omega\mu J_z - \frac{j}{\omega\epsilon}\frac{\partial}{\partial z}(\text{div}\bar{J}) \tag{9}$$

obtained from Maxwell's equations. Let E_z, E_{zp} and the known right-hand member of (9) be expanded in terms of the complete set $[E_{zn}]$, and the expansions inserted in (9). If the coefficients of $[E_{zn}]$ in both members are equated, using (6), differential equations for the z dependent coefficients of the expansions result. These equations lead to the following:

$$E_z = E_{zp} + \sum_n A_n e^{jh_n z}[E_{zn}] + \sum_n B_n e^{-jh_n z}[E_{zn}]. \tag{10}$$

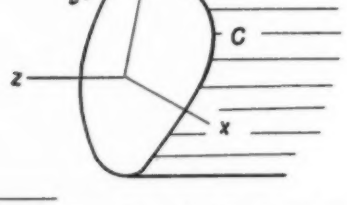


FIG. 1.

* Department of Electrical Engineering. Graduate fellow of the Belgian American Educational Foundation. Now with Manufacture Belge de Lampes Electriques S.A. (Mazda), Brussels, Belgium.

[1] J. C. Slater and N. Frank, *Electromagnetism* (McGraw-Hill Book Company, Inc., New York, 1947), Chapter XI.

[2] H. Poincare, Rendic. Circ. Math. di Palermo **VIII**, 57 (1894).

The terms of the summations are recognized from (1) as being the E_z components of the normal modes. Consolidating in one symbol the two waves into which each mode degenerates (one directed toward negative z and one toward positive z), one can write, in short,

$$E_z = E_{zp} + \sum_n C_n E_{zn}. \tag{11}$$

Similarly it can be established that

$$H_z = H_{zp} + \sum_m D_m H_{zm}. \tag{12}$$

Equations (11) and (12) show how the z components of the field can be expanded in the z components of the normal modes. It is now desired to extend that expansion to the complete vectors, using the same expansion coefficients C_n, D_m.

First, let the dielectric region be simply connected. It is desired to show that the following equations must hold:

$$\bar{E} = \bar{E}_p + \sum_n C_n \bar{E}_n{}^E + \sum_m D_m \bar{E}_m{}^H, \tag{13}$$

$$\bar{H} = \bar{H}_p + \sum_n C_n \bar{H}_n{}^E + \sum_m D_m \bar{H}_m{}^H. \tag{14}$$

From (11) and (12) let it be seen that the field distributions described by left and right members of (13) and (14) have the same z components. They satisfy identical Maxwell's equations and boundary conditions, the sources $\bar{J}$ being identical for $\bar{E}$, $\bar{H}$ and $\bar{E}_p$, $\bar{H}_p$. To show that these field distributions are equal, it suffices to prove that their difference $\bar{E}_d$, $\bar{H}_d$ is zero. The fields $\bar{E}_d$, $\bar{H}_d$ have no z components, and satisfy Maxwell's equations in free space, the sources $\bar{J}$ disappearing by subtraction. This reduces the proof to an electrostatic problem. Indeed, from (2) projected on the z axis and (5), it is easily established that $\bar{E}_d$ satisfies Laplace's equation (8). It also satisfies from (2) and (3),

$$\nabla^2 \bar{E}_d + (\omega^2/c^2)\bar{E}_d = 0, \tag{15}$$

so that

$$\frac{\partial^2 \bar{E}_d}{\partial z^2} + \frac{\omega^2}{c^2}\bar{E}_d = 0, \tag{16}$$

and

$$\bar{E}_d = \bar{F}_+(x, y)e^{j(\omega/c)z} + \bar{F}_-(x, y)e^{-j(\omega/c)z}. \tag{17}$$

It is now possible to show that $\bar{F}_+ = \bar{F}_- = 0$. For, having $\nabla_{xy}{}^2 \bar{E}_d = \mathrm{div}\bar{E}_d = (\mathrm{curl}\bar{E}_d)_z = (\bar{E}_d)_t = 0$ for all the values of z between S_1 and S_2, it follows by (17) that $\nabla^2 \bar{F}_\pm = \mathrm{div}\bar{F}_\pm = \mathrm{curl}\bar{F}_\pm = (\bar{F})_t = 0$. In conclusion, $\bar{F}_\pm$ satisfy the same equations and boundary conditions as would the electrostatic field inside C in the absence of charges. This field vanishes in a simply connected dielectric region. Thus $\bar{F}_\pm$ are both zero, as are $\bar{E}_d$(17) and $\bar{H}_d$(2).

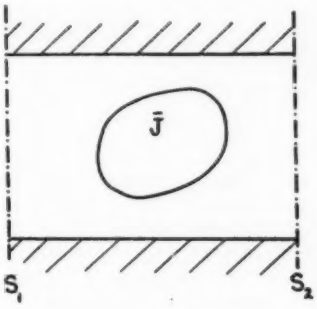


FIG. 2.

When the dielectric is multiply connected, an electrostatic field $\bar{F}_0(x, y)$ can exist and from (17)

$$\bar{E}_d = \bar{F}_0[A_0 e^{j(\omega/c)z} + B_0 e^{-j(\omega/c)z}] = A_0\bar{E}_0{}^+ + B_0\bar{E}_0{}^- = C_0\bar{E}_0, \tag{18}$$

where $\bar{E}_0{}^+$ and $\bar{E}_0{}^-$ are precisely the two waves which comprise the *TEM* mode already mentioned above. Substituting then, for (9) and (10)

$$\bar{E} = \bar{E}_p + \sum_n C_n \bar{E}_n + \sum_m D_m \bar{E}_m + C_0\bar{E}_0, \tag{19}$$

$$\bar{H} = \bar{H}_p + \sum_n C_n \bar{H}_n + \sum_m D_m \bar{H}_m + C_0\bar{H}_0, \tag{20}$$

the following conclusion hold for both cases:

(a) Between two cross sections of a guide which enclose no currents, the field can be split into a sum of normal modes, with coefficients determined by the sources outside the volume (S_1, S_2).

(b) Between two cross sections which enclose currents, the expansion must be completed by an extra term $\bar{E}_p\bar{H}_p$. This term depends only upon the enclosed currents, and is independent of conditions outside the volume (S_1, S_2), namely, of current sources external to it, and of the way in which the guide is coupled with the outside space at its terminals.

Certain Refractory Compounds as Thermionic Emitters

D. L. GOLDWATER AND R. E. HADDAD
Bartol Research Foundation of the Franklin Institute, Swarthmore, Pennsylvania
(Received July 17, 1950)

Several refractories of the nitride, boride, and carbide families are investigated as sources of thermionic electrons. An activation pattern qualitatively common to several of the compounds is observed. Emission data for the final state of activation are presented. Data concurrently obtained for thoria are also given for comparison.

True temperature used in the Richardson equation is obtained by correction from brightness temperature. Temperature correction curves have been obtained. These curves can be satisfactorily represented in the region of thermionic emission by one value of spectral emissivity, which values are listed for the materials studied.

I. INTRODUCTION

A SURVEY of refractory compounds as thermionic emitters is in progress in this laboratory. The materials which have been investigated thus far have been several carbides, borides, and nitrides which were immediately available from a chemical supplier.* The primary purpose of such a program is to uncover some material which shows superiority in some respect to emitters now commonly in use. Of all the materials tested, encouraging results were obtained for only one, zirconium carbide, and more detailed tests of its usefulness as an electron emitter are planned.

II. PREPARATION OF EXPERIMENTAL TUBES

The materials tested were: Zirconium carbide (ZrC), thorium carbide (ThC_2), titanium carbide (TiC), and tantalum carbide (TaC); zirconium nitride (ZrN), titanium nitride (TiN), and tantalum nitride (TaN); zirconium boride (ZrB) and tantalum boride (TaB).† In most cases these were used as supplied. In certain cases the majority of the particles were too large to form a satisfactory coating. These were ball-milled, then washed in concentrated HCl to remove the iron introduced during ball-milling.

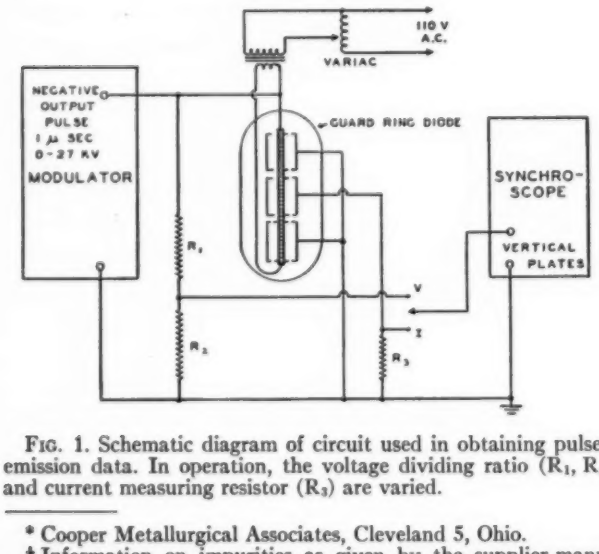


FIG. 1. Schematic diagram of circuit used in obtaining pulsed emission data. In operation, the voltage dividing ratio (R_1, R_2) and current measuring resistor (R_3) are varied.

The materials were used as coatings on 12-mil tungsten wire. These filaments were preprocessed by heating to the region of 1800°C in hydrogen, while under tension. They were thus simultaneously cleaned and straightened.

All the nitrides and zirconium carbide were applied to the wire by electrophoresis. The material to be used was suspended in butyl alcohol, with a trace of electrolyte added. The filament was placed in the suspension, and made about 100 v negative with respect to a cylindrical anode surrounding it. After a few seconds, the filament was withdrawn. Control of thickness is obtainable to some extent through the length of time the filament is held in the suspension. This method can be made to yield very uniform coatings. The other materials would not readily form coatings by the aforementioned method. They were painted on the filament, titanium carbide in butyl alcohol suspension, and all others in nitrocellulose binder.

Finished coatings were examined under the microscope for uniformity, and measured at several points by means of a micrometer eyepiece. The mean diameter was used in the calculation of the emitting area. Long filaments, about 5 cm, were used, and at least 3 cm at the center, coated. Thus a uniform emitting surface was present extending well beyond the limits of the center anode of the guard-ring system. The cathodes thus prepared were mounted in conventional guard-ringed diodes, the length of the central anode being 1 cm.

III. PROCESSING AND TESTING OF EXPERIMENTAL TUBES

Tubes were evacuated on either an oil or a mercury pump; no differences which could be attributed to pumping medium were observed. Processing was always initiated by a 450°C bake of at least 1 hr. Pressure after the bake was always lower than 2×10^{-7} mm of Hg. The bake was followed by intense inductive heating of the anode system, to about 1250°C. Subsequent heat-

* Cooper Metallurgical Associates, Cleveland 5, Ohio.

† Information on impurities as given by the supplier-manufacturer: The borides all contain traces of Fe, Ni, Si, and C. In the case of TaB, the compound TaB_2 may be present to the extent of several percent. ThC_2 has Si impurity; TaC contains Cb and Ti. ZrN may contain sizable impurity of Zr_3N_2.

ing by electron bombardment was never allowed to raise the temperature above 1100°C. After anode outgassing, the filament was raised quickly to incandescence. With the temperature around 900°C anode voltage was applied. Customary procedure was to take data for a Schottky determination of I_s at the temperature at which saturation currents of 100–200μa/cm^2 were first available. The temperature was then raised in steps to 1500°C or higher, with several emission runs taken at intervals of 100°–200°C.

Saturation currents were limited by anode dissipation to 60–80 ma. Beyond this level no dc data were taken. The cathodes were, in general, outgassed at 1800°–2000°C, as subsequent pulse data were taken at these higher temperatures. Before seal-off, cathodes were seasoned at some temperature in the 1500–1700 degree region, with sizeable dc being drawn. The effect

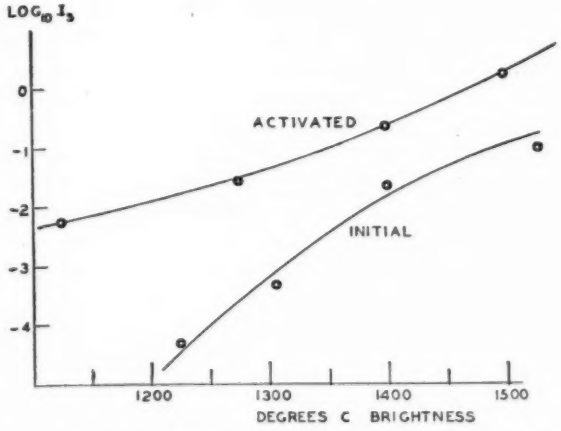


FIG. 2. The "initial" emission is that obtained on reaching a given temperature during the first heating of the cathode. The "activated" data are for the reproducible state of final activation. These data were obtained during the processing of one ZrC cathode.

of this seasoning will be described and illustrated subsequently.

After seal-off, pulsed and dc Schottky data were obtained, and plots of $\log I_s/T^2$ *vs* $1/T$ constructed. The pulsed data were taken using a Link model 4 modulator and an MIT Radiation Laboratory synchroscope, in the circuit of Fig. 1. A low duty cycle of 0.00006 permitted the use of high peak voltage and current without noticeable anode heating. It will be noted that it was possible to overlap the two methods of measurement somewhat; the emission parameters, A and ϕ, indicated by the two groups of data considered separately are close to those finally determined from the combined data.

IV. TEMPERATURE MEASUREMENT AND CORRECTION

Two conditions are necessary in order that the temperature of the coated filament, as measured with a pyrometer, be useful in determining the emission constants. The first is that this temperature be uniform over the area from which measured emission is drawn. Emission measurements were considered acceptable when the variation along the central portion of the cathode was not greater than $\pm\frac{1}{2}$ percent of the given temperature.

TABLE I.

Material	Spectral emissivity at 0.655 micron
ThC_2	Approx. 0.8
ZrC	0.96
TiC	0.96
TaC	0.67
TaB	0.70
ZrB	0.70
ThO_2	0.40

It is also necessary to know the spectral emissivity of the coating material. Spectral emissivity measurements have been made.[1] For every material except thoria, the single value of emissivity given in Table I relates true to brightness temperature over the whole experimental range. In the case of thoria, the value is good to about 1600°C (brightness temperature), above which point the emissivity increases towards unity.

V. RESULTS

A. Activation During Processing

It has been mentioned that a period of seasoning the cathode preceded the seal-off. Marked improvement of emission usually occurred during this treatment. The seasoning period usually included continuous drawing of current. At the beginning of the period this was always emission limited, but frequently the cathode would improve to such an extent that it would be necessary to reduce the anode voltage at intervals in order that the anode would not overheat. Figure 2 shows the effect of 70 min of seasoning at 1500°C on a particular zirconium carbide cathode, giving data taken before and after seasoning.

TABLE II.

Material	ϕ Volts	A Amp cm^2	I_s −1800°K Amp cm^2	I_s −2000°K Amp cm^2 deg K^2
	(a)			
Thoria	2.55	3.3	2.6	11
Zirconium carbide	2.18	0.31	0.9	4.0
Titanium carbide	2.35	25	0.03	0.30
Tantalum carbide	3.14	0.30	0.04	0.32
Tantalum boride	2.89	10	0.31	2.3
	(b)			
Zirconium boride	(4.48)	(35000)	0.03	0.6
	(c)			
Thorium carbide	(3.5)	(550)	0.35	(4)
Zirconium nitride	—	—	—	0.05
Boron nitride	—	—	—	0.04
Titanium nitride	—	—	—	0.05

[1] F. H. Morgan, J. Appl. Phys. **22**, 108 (1951).

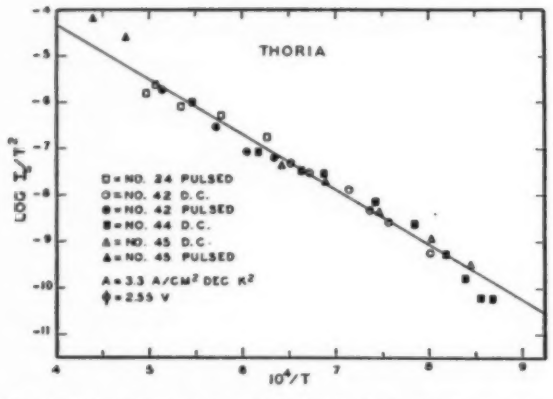


FIG. 3. The values of A and ϕ obtained for thoria are in good agreement with those already obtained by D. A. Wright, T. E. Hanley, and O. A. Weinreich.

It is possible to plot the temperature-emission data taken during the initial heating (i.e., before seasoning) on scales of $\log I_s/T^2$ *vs* $1/T$. The points which comprise such a plot, obtained during the initial raising of the temperature, cannot in general be reproduced at any subsequent time in the history of the cathode. It is believed that each such point represents a different state of partial activation of the cathode. The Richardson equation,

$$I_s = AT^2 \exp(\phi e/KT),$$

when applied to such data, must be regarded merely as a useful analytical function which happens to fit the data fairly well. With these reservations, it is noted that the values of A and ϕ yielded by the pre-activation data are extraordinarily high; in one case an A of 10^6 amp/cm²deg² has been observed.

The final activation state after seasoning was deemed to be stable when no further improvement occurred at the seasoning temperature. Several factors justify describing this state as stable. First, after seasoning, flashing to higher temperatures did not produce any lasting improvement in the emission. Second, emission data taken over the whole temperature range (in the case of zirconium carbide the range is from 900° to 1800°C) were consistent and reproducible. Third, current drawing did not affect the emission, except, in some cases, high dc densities. In these last cases, the emission "decays" with time. In the stable emission state, however, this was found to be reversible when the cathode was allowed to recover at the same temperature, but with no emission current.

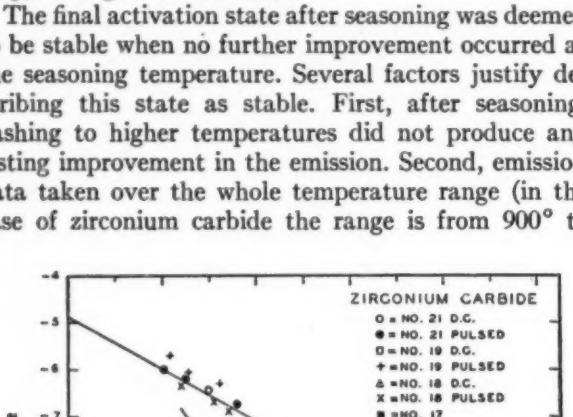


FIG. 4. The dotted line represents data taken before activation on one tube; i.e., the "initial" data of Fig. 2.

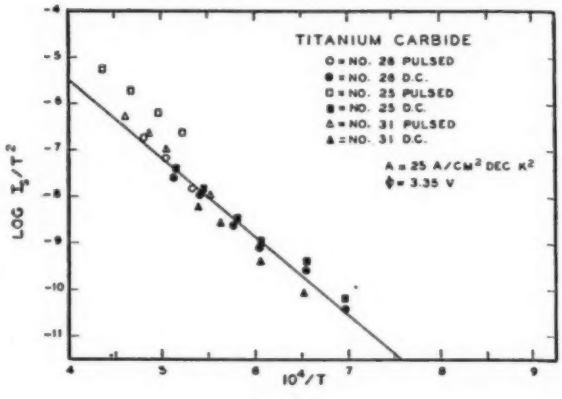


FIG. 5.

In regard to the effect of flashing, it should be mentioned that Weinreich[2] has observed "states of higher activation" for thoria. This enhanced thermionic emission may be induced by flashing to very high temperature (order of 3000°K). The very high emission available from this state is rapidly reduced by current drawing, or by the presence of gases (e.g., 10^6 mm of oxygen). Once gone, it does not recover in the manner of the decay previously mentioned. Our work does not exclude the possibility of such a transient state of high activation for the materials examined.

B. Values for Thermionic Emission, and for the Parameters A and ϕ

Five of the materials tested (including thoria) adhered closely to the pattern of activation described in the preceding section. Table IIa gives the values of the

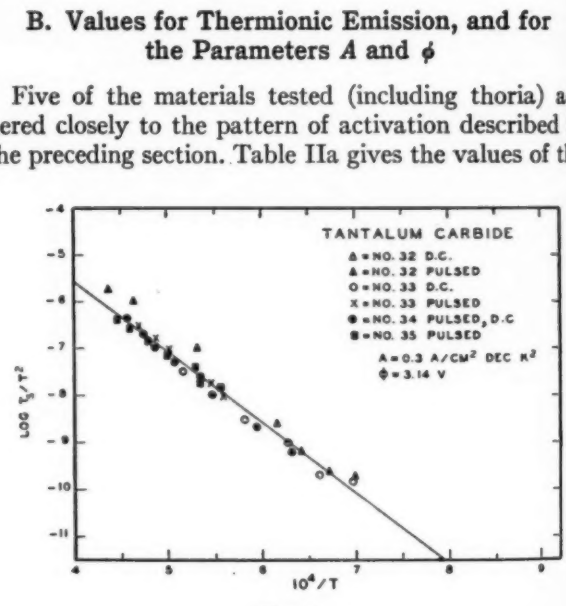


FIG. 6.

[2] O. A. Weinreich, J. Appl. Phys. **20**, 1256 (1949); O. A. Weinreich, Rev. Gén. d'El. **54**, 243–256 (1945).

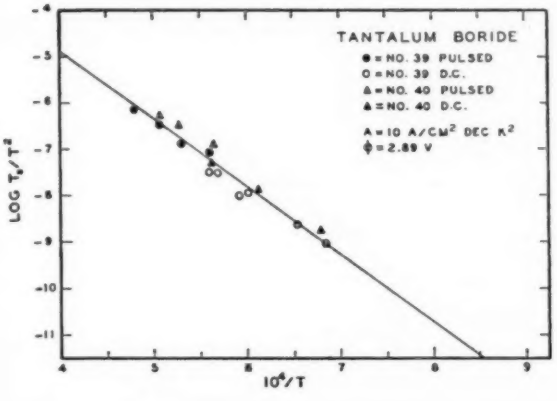


FIG. 7.

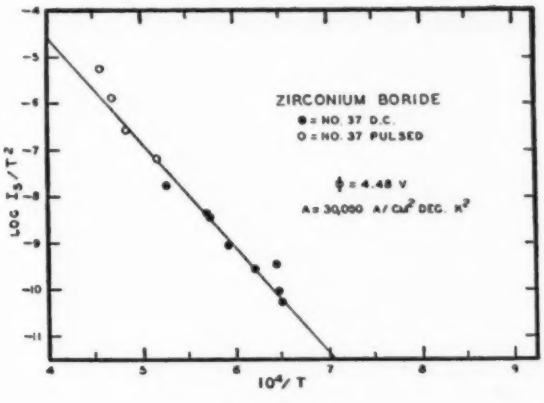


FIG. 8.

emission parameters for these materials, as determined from the data of Figs. 3 through 7. The last two columns of this table show emission values for two temperatures. It is seen that thoria is a considerably better emitter, in this temperature range, than any of the new materials tested. (This statement will be qualified in the concluding section, in which the practical possibilities of ZrC are discussed.)

Table IIb and Fig. 8 show data for a material which did not adhere to the activation pattern. During seasoning at 1500°C, operation at 1700°C and flashing to considerably higher temperatures, zirconium boride appeared to maintain a stable state of low activation.† The data represent only one experimental tube.

In the case of thorium carbide (Table IIc), the coating appeared to flake and fall off during activation. The problem appears to be thermomechanical; there was no evidence of evaporation, and it is expected that by adjustment of coating technique cathodes can be prepared which will yield more complete data. There was evidence of considerable activation; emission of about 0.3 amp/cm² at 1500°C was obtained with one tube. The figures given probably represent a state of partial activation which prevailed when the cathode failed for the reason mentioned. It is possible that further experiments with more satisfactory coatings may show thorium carbide to lie in the range of practical usefulness.

All three nitrides tested (Table IIc) evaporated rapidly at temperatures which yielded very low emission. A typical diode gave emission of 50 ma/cm² at 1700°C. With no further increase of heater power, three effects were observed, all attributable to rapid evaporation of the coating: (1) Increase in cathode temperature (as coating evaporates, total thermal radiation decreases, and with constant power, temperature increases); (2) tendency for gas discharge to occur due to high vapor pressure; and (3) rapid blackening of the envelope.

C. Practical Possibilities of ZrC

It will be seen in Table IIa that ZrC is the best thermionic emitter of the new materials tested. In the region of practical usefulness, the emission available is less than 40 percent of that which is expected of thoria. However, the decay of emission usually observed with thoria and conventional oxide cathodes, when high density emission currents are continuously drawn, was not observed in the case of zirconium carbide. Thus, up to the limitation imposed by the anode dissipation in the test diode (about 0.8 amp/cm²), the dc tests showed the same emission current (at a given temperature and anode voltage) as the pulse tests. Two of the ZrC cathodes were operated for several hours, at 1500°C, with continuous space charge limited current of 0.7 amp/cm² drawn. One would expect to obtain higher current densities at higher cathode temperatures, if the anode did not limit the tube performance.

It must be noted, however, that no ZrC cathode has been tested for life beyond 30 hr. High current density life-tests of zirconium carbide cathodes will be undertaken. Ability to supply high density of emission continuously is of great importance in many applications, and zirconium carbide cathodes may well prove to be of interest in this respect.

† It has already been observed that abnormally high values of A and θ are generally found when we represent the data of the poorly activated states by the Richardson equation.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Calculation of Diffusion Penetration Curves for Surface and Grain Boundary Diffusion

J. C. FISHER
General Electric Research Laboratory, Schenectady, New York
(Received July 19, 1950)

Diffusion in solids is known to occur along grain boundaries and over free surfaces more rapidly than through the interiors of crystals. In order to facilitate quantitative investigation of grain boundary and surface diffusion, a mathematical analysis of the problem has been completed, assuming that grain boundary diffusion is analogous to the diffusion of heat along a thin copper foil imbedded in cork. The calculated diffusion-penetration relationship for grain boundary diffusion is shown to agree with the experimentally determined grain boundary self-diffusion of silver.

THE self-diffusion of silver has been studied in detail during the past year.[1] It has been shown that, at low temperatures, penetration of radioactive silver into polycrystalline silver takes place much more rapidly than does penetration into single crystals. At higher temperatures, the rates of penetration approach one another and very nearly become equal. The high rate of penetration in polycrystals at low temperatures is associated with grain boundary diffusion, an understanding of which is essential for satisfactory interpretation of the low temperature diffusion process.

There are three types of paths along which diffusion can occur in crystalline solids: volume diffusion through the interiors of crystals, grain boundary diffusion along the disordered internal interfaces of polycrystals, and surface diffusion over single crystals or polycrystals.

Volume self-diffusion follows the law,

$$P = -D \operatorname{grad} C, \tag{1}$$

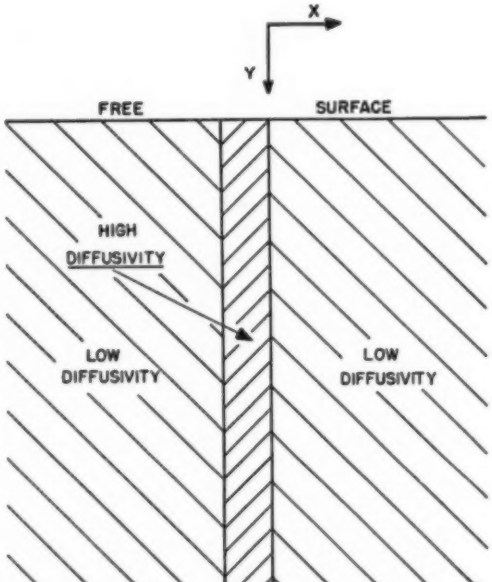


FIG. 1. Coordinate system for diffusion outside a semi-infinite slab of high diffusivity material imbedded in a semi-infinite solid.

where P is the flux and C the concentration of diffusing material, and D is constant. The dependence of the diffusion constant upon temperature is very well represented by a relationship of the form,

$$D = D_0 \exp(-Q/RT), \tag{2}$$

where Q is the activation energy for diffusion.

Assuming that grain boundary and surface diffusion coefficients vary with temperature according to Eq. (2), Langmuir[2] estimated activation energies for the diffusion of thorium in and on wolfram,

$$Q_v = 120 \text{ kcal/g atom},$$
$$Q_{gb} = 90 \text{ kcal/g atom},$$
$$Q_s = 66 \text{ kcal/g atom},$$

where Q_v, Q_{gb}, Q_s refer to volume, grain boundary, and surface diffusion.

Accurate determinations of Q_{gb} and Q_s, and particularly of the coefficients D_0 for grain boundary and surface diffusion, require a more detailed analysis of surface and interface diffusion than hitherto has been available. An improved analysis is particularly important for the study of self-diffusion, where adsorption does not tend to confine diffusing material to the interface.

An intuitive picture of grain boundary diffusion is that of diffusion along a thin layer of high diffusivity material sandwiched between large volumes of low diffusivity material, in a manner analogous to the diffusion of heat along a copper foil imbedded in cork. A similar picture suggests itself for surface diffusion. The consequences of adopting this intuitive picture will be examined, and the resulting calculated diffusion penetration curves will be compared with experimental curves.

It is necessary to consider only grain boundary diffusion, for surface diffusion is mathematically identical. The problem is to determine the concentration of diffusing material in a semi-infinite solid having a semi-infinite slab of highly permeable material imbedded in it, as shown in Fig. 1, the plane of the slab being normal to the free surface. The desired boundary condition is a zero initial concentration in the solid, with the surface maintained at a fixed concentration from zero time onward.

[1] R. E. Hoffman and D. Turnbull, "Lattice and grain boundary self-diffusion in silver" (to be published).

[2] I. Langmuir, J. Franklin Inst. **217**, 543 (1934).

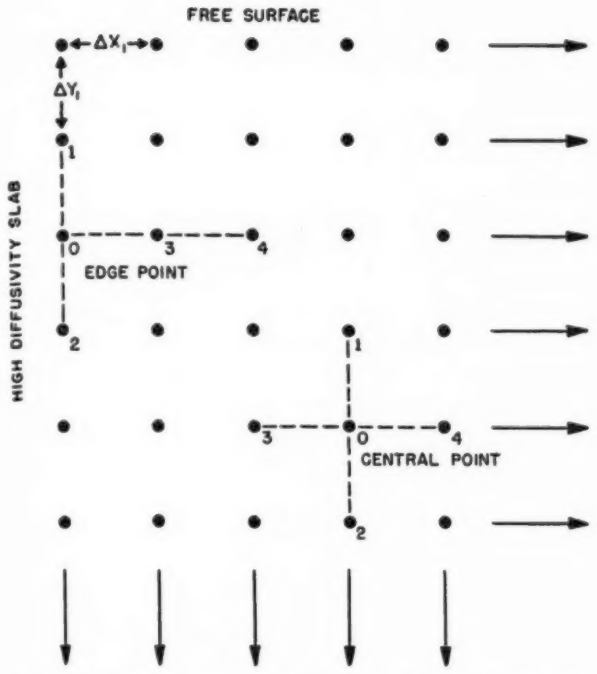


FIG. 2. Network for numerical solution of diffusion equations.

Let the thickness of the slab be δ. Take the y axis normal to the free surface and take the x axis normal to the slab. It is assumed that the slab is so thin that concentration variations across its width are negligible. Effectively, the instantaneous concentration in the slab is a function of y only. Concentration in the slab then varies according to the equation

$$\partial C/\partial t = (D'\partial^2 C/\partial y^2) + (2/\delta)D(\partial C/\partial x)_0, \qquad (3)$$

where the first term represents the concentration change associated with diffusion of material along the slab, and the second term (which should be divided by two for surface diffusion) represents the change associated with loss of material from the two sides of the slab. D' and D are the diffusion coefficients for diffusion in and outside the slab, and the subscript zero on $\partial C/\partial x$ means that the derivative is to be evaluated just outside the slab, at its surface. The concentration outside the slab varies according to the equation

$$\partial C/\partial t = D\nabla^2 C. \qquad (4)$$

Whenever the penetration of diffusing material is much greater in the slab than elsewhere, diffusion outside the slab is primarily normal to the slab (except in a thin layer of material at the free surface of the solid, where it is normal to the free surface) and the right-hand term of Eq. (4) becomes substantially $D\partial^2 C/\partial x^2$. Equations (3) and (4) now can be put in a more convenient form by multiplying by δ^2/D and making the substitutions,

$$\begin{aligned} t_1 &= Dt/\delta^2, \\ x_1 &= x/\delta, \\ y_1 &= y/[\delta(D'/D)^{\frac{1}{2}}]. \end{aligned} \qquad (5)$$

Equation 3 becomes

$$\partial C/\partial t_1 = (\partial^2 C/\partial y_1^2) + 2(\partial C/\partial x_1)_0, \qquad (3a)$$

and Eq. (4) becomes

$$\partial C/\partial t_1 = \partial^2 C/\partial x_1^2. \qquad (4a)$$

Equations (3a) and (4a) have been solved by two approximate methods. The first, a numerical method, is based upon the difference equations

$$\Delta C/\Delta t_1 = (C_1 - 2C_0 + C_2)(1/\Delta y_1)^2 + 2(-1.5C_0 + 2C_3 - 0.5C_4)(1/\Delta x_1), \qquad (3b)$$

in the slab, and

$$\Delta C/\Delta t_1 = (C_3 - 2C_0 + C_4)(1/\Delta x_1)^2, \qquad (4b)$$

outside the slab, which approximate Eqs. (3a) and (4a). Figure 2 shows the locations of the appropriate concentrations C_i for computing ΔC at an edge point [Eq. (3b)] and at a central point [Eq. (4b)]. ΔC is the concentration increment at point 0 corresponding to the time increment Δt_1. A separate numerical solution is required for each pair of D'/D, δ values. Figure 3 (upper values) gives a solution of Eqs. (3b) and (4b) for time $t_1 = 100$, and for the particular diffusion coefficient ratio

$$D'/D = 10^6,$$

and the particular grain boundary thickness

$$\delta = 5(10)^{-8} \text{ cm}.$$

	$X_1=0$	$X_1=16$	$X_1=32$	$X_1=48$	$X_1=64$
$y_1=0$	1000 1000	239 258	31 24	2 1	0 0
$y_1=4$	245 261	45 67	5 6	0 0	
$y_1=8$	58 68	9 18	1 2	0 0	
$y_1=12$	18 18	2 5	0 0		
$y_1=16$	3 5	0 1	0 0		
$y_1=20$	0 1	0 0			
$y_1=24$	0 0				

FIG. 3. Concentration of diffusing material as a function of position. Upper values—numerical solution of Eqs. (36), (46). Lower values—Eq. (8). $t_1 = 100$, $D'/D = 10^6$, $\delta = 5(10)^{-8}$ cm.

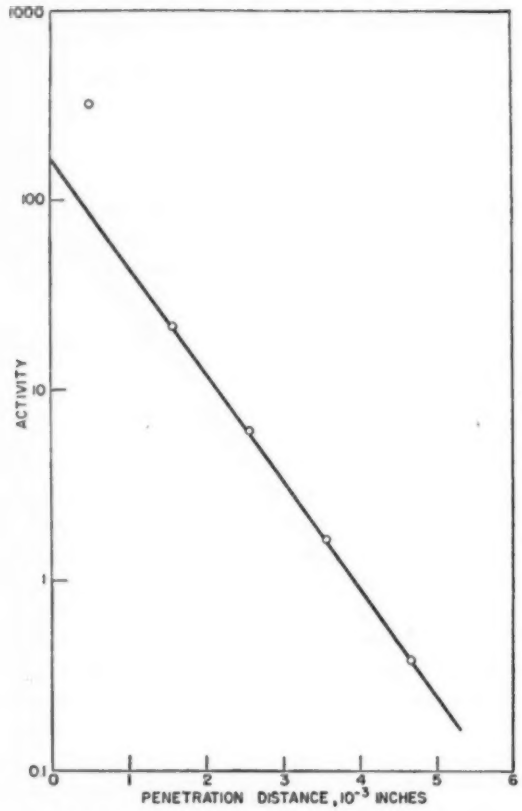


FIG. 4. Activity (log scale) *versus* penetration distance for grain boundary diffusion of radioactive silver into polycrystalline silver, 63 hours at 478C.

It is estimated that the accuracy of the C-values is about ± 10 percent or $\pm 0.005\, C_{max}$, whichever is larger, where C_{max} is the value of C at the free surface.

The second approximate solution of Eqs. (3a) and (4a) is an analytical solution suggested by certain observations regarding the numerical solution. It was observed that C-values in the slab rise at a rapidly decreasing rate, thereby spending most of their time near their current values. Hence, with respect to diffusion outside the slab, it can be assumed, with little error, that the concentration in the slab has had the distribution characteristic of time $t=t$ since time $t=0$. In other words, if $\phi=\phi(y_1)$ is the correct concentration in the slab at time t_1, the concentration outside the slab is approximately

$$C=\phi \operatorname{erfc}(x_1/2t_1^{\frac{1}{2}}), \tag{6}$$

where

$$C=C' \operatorname{erfc}(x_1/2t_1^{\frac{1}{2}}),$$

is the exact concentration outside the slab corresponding to a constant concentration C' in the slab.* It was further observed that most of the time $\partial C/\partial t$ in the slab $\gg (\partial C/\partial t)_{max}$ outside the slab, suggesting that the true function $\phi(y_1)$ is not significantly different from that for which $\partial C/\partial t_1=0$ in the slab. The condition $\partial C/\partial t_1=0$ applied to Eq. (6) requires that ϕ satisfy the equation

$$(\partial^2\phi/\partial y_1^2)-(2\phi/\pi^{\frac{1}{2}}t_1^{\frac{1}{2}})=0. \tag{7}$$

The solution, with the boundary condition $C=1$ at $y_1=0$, is

$$\phi=\exp(-2^{\frac{1}{2}}y_1/\pi^{\frac{1}{2}}t_1^{\frac{1}{2}}),$$

giving as the final approximate expression for the concentration

$$C=\exp(-2^{\frac{1}{2}}y_1/\pi^{\frac{1}{2}}t_1^{\frac{1}{2}})\operatorname{erfc}(x_1/2t_1^{\frac{1}{2}}). \tag{8}$$

Values of C calculated from Eq. (8) are compared in Fig. 3 with those calculated from Eqs. (3b) and (4b). The disagreement between the two methods of calculation does not exceed the estimated error of the numerical analysis, supporting the view that Eq. (8) is a reasonable approximation to the true dependence of concentration upon x_1, y_1, and t_1.

It is of interest to compare the predictions of Eq. (8) with the results of experiment. Consider the diffusion of radioactive silver into a polycrystalline silver specimen from a thin electrodeposited layer of radioactive metal. At low temperatures, most penetration should be via grain boundary diffusion, along grain boundaries the majority of which are nearly at right angles to the specimen surface. If, as seems not unreasonable, surface diffusion and diffusion in the highly disordered electrodeposited layer of radioactive silver are sufficiently rapid to maintain a nearly steady concentration of radioactive silver at the exposed edge of each grain boundary, the distribution of radioactive silver at each grain boundary should be given by Eq. (8).

The usual experimental observation concerning diffusion into a semi-infinite solid is expressed by means of a diffusion penetration curve, which gives the quantity of diffusing material that has penetrated the solid as a function of the penetration distance. Whenever the penetration distance corresponding to bulk diffusion is significantly less than that corresponding to grain boundary diffusion, a theoretical diffusion penetration curve can be calculated from Eq. (8), according to which the quantity of diffusing material lying between parallel planes at y_1 and at y_1+dy_1 in the semi-infinite solid is

$$Q=\int_V C dV=\int_{y_1}^{y_1+dy_1}\int_{-\infty}^{\infty} C dx_1 dy_1. \tag{9}$$

Performing the integration,

$$Q\sim t_1^{\frac{1}{2}}\exp(-2^{\frac{1}{2}}y_1/\pi^{\frac{1}{2}}t_1^{\frac{1}{2}})dy_1, \tag{10}$$

where a precise value for the constant of proportionality is not required.

It is evident that the radioactivity A of a sample of silver machined from between depths $y_1-\Delta y_1/2$ and $y_1+\Delta y_1/2$ in the sample will vary with y_1, t_1, and Δy_1 according to the proportionality

$$A\sim t_1^{\frac{1}{2}}\exp(-2^{\frac{1}{2}}y_1/\pi^{\frac{1}{2}}t_1^{\frac{1}{2}})\Delta y_1, \tag{11}$$

as long as the value of Δy_1 is not too large.

* $\operatorname{erfc} x=1-\operatorname{erf} x=(4/\pi)^{\frac{1}{2}}\int_x^\infty \exp(-y^2)dy$.

The specific activity, $a=A/\Delta y_1$, is

$$a\sim t_1^{\frac{1}{2}}\exp(-2^{\frac{1}{2}}y_1/\pi^{\frac{1}{4}}t_1^{\frac{1}{4}}). \quad (12)$$

Taking logarithms,

$$\log a=-(2^{\frac{1}{2}}\log e/\pi^{\frac{1}{4}}t_1^{\frac{1}{4}})y_1+\text{const.},$$

or

$$\log a=-[2^{\frac{1}{2}}(D/D'\delta)^{\frac{1}{2}}\log e/(\pi Dt)^{\frac{1}{4}}]y+\text{const.}, \quad (13)$$

where the constant depends upon t, but not upon y. According to Eq. (13), a plot of log activity *versus* depth of penetration should yield a straight line of slope

$$d\log a/dy=-2^{\frac{1}{2}}(D/D'\delta)^{\frac{1}{2}}\log e/(\pi Dt)^{\frac{1}{4}}. \quad (14)$$

In Fig. 4, experimentally determined activity values for diffusion of radioactive silver into polycrystalline silver (63 hours at 478C), as determined by Hoffman and Turnbull,[1] are plotted on a logarithmic scale *versus* depth of penetration. As predicted by Eq. (13), the points lie upon a straight line, within the limit of experimental error. The point nearest the free surface lies well above the straight line, as it should, for it represents the activity of the first slice of silver machined from the specimen, including the activity of the electrodeposited surface layer. According to Eq. (14) the slope of the line, together with the known value of D for bulk diffusion at 478C[1] and the known diffusion time t, fixes the value of $(D/D'\delta)$ at

$$D/D'\delta=21.3\ \text{cm}^{-1}. \quad (15)$$

In order to obtain a value for D' or for the ratio D'/D, a grain boundary thickness must be assumed. It is not unreasonable to assume a thickness of about two atom layers,

$$\delta\approx 5(10)^{-8}\ \text{cm}. \quad (17)$$

With this value of δ,

$$D'/D=0.94(10)^{6}, \quad (18)$$

is obtained for the ratio of the grain boundary diffusion coefficient to the bulk diffusion coefficient at 478C.

It is concluded that diffusion penetration curves for grain boundary self-diffusion are adequately approximated by Eq. (13), and that the corresponding concentration of diffusing material in the neighborhood of a grain boundary is adequately approximated by Eq. (8).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Secondary Emission of Nickel-Barium Mixtures and Rhenium When Bombarded by Electrons with Energies from 50 to 8000 Electron-Volts*

H. E. FARNSWORTH
Brown University, Providence, Rhode Island
AND
M. J. LUN
Philips Laboratories, Inc., Irvington, New York

(Received July 28, 1950)

The experimental tube contains a contamination-free gun which directs an electron beam at normal incidence onto a small thin sheet target which may be heated by conduction. The target is located at the center of a conducting sphere to permit a measure of the secondary current, and also of the energy distribution of the secondary electrons by means of the retarding potential method.

The maximum value of δ, the secondary emission ratio, for a cast alloy of nickel containing 1.5 percent barium is 2.8 and occurs at 800 to 900 volts after a critical amount of previous heat treatment of target. The value of δ_{max} increases with heat treatment of target, subsequent to baking of tube, to the above value and then decreases with further heating. For the critical condition the value of δ above 4000 volts is sensitive to the amount of previous electron bombardment of the target. A δ_{max} of 1.3 for rhenium after thorough outgassing was obtained at 900 volts.

The most probable energy of low speed secondary electrons from rhenium is approximately 5 ev.

INTRODUCTION

THE present work on nickel +1.5 percent barium and on rhenium was undertaken in conjunction with a larger investigation of these and other materials for thermionic emission properties. Although many composite and pure emitters have been investigated, there has been found no single secondary emission surface which fulfills the stringent requirements necessary in the operation of high powered magnetrons. It is the purpose of these experiments to obtain basic information on the nickel-barium alloy to determine if it displays properties which might be useful in c.w. magnetron design. In past work on pure metals, no reference was found on the secondary emission ratio for rhenium. Hence measurements for this element are included in this study. Results for nickel, barium, and barium oxide[1,2] are available over limited voltage ranges, so that

* Partial fulfillment of ONR Contract No. N6onr-261 Task Order I. This work was done at the Philips Laboratories.

[1] H. Bruining, *Die Sekundär-Elektronen-Emission fester Körper* (Verlag. Julius Springer, Berlin, 1942). K. G. McKay, "Secondary Electron Emission", *Advances in Electronics* (Academic Press, Inc., New York, 1948), Vol. 1.

[2] J. B. Johnson, Phys. Rev. 73, 1058 (1948).

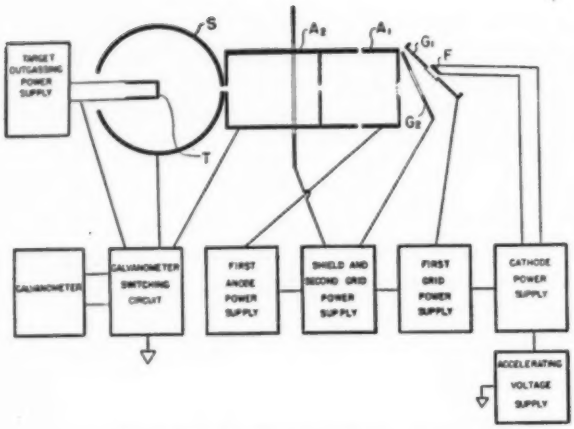


FIG. 1. Block diagram of apparatus.

comparisons may be made with the results presented here for the Ni-Ba alloy.

APPARATUS AND PROCEDURE

The experimental tube has a gun structure which collimates a beam of electrons that strikes the target at normal incidence. The filament cathode, F, Fig. 1, consists of a flat tantalum strip which is heated by conduction. This is mounted off the gun axis to prevent evaporated atoms from striking the target. The metal parts are constructed of tantalum. Fine tantalum gauze covers the 3 mm holes in G_1 and A_1. The current leads support the target T through an opening in the back of the sphere S. The metal parts are held in place by cementing metal supports to fused quartz rods as shown in Fig. 2. This type of support is required to prevent electrical leakage at the high voltages used. The tube envelope is of Nonex glass with leads spaced so as to prevent electrical leakage. A conducting guard ring on the inside of the tube supporting the sphere connection also prevents leakage. The electrons are accelerated to G_1, deflected in the electrostatic field produced by a negative voltage on the arc-shaped grid G_2,[3] then pass along the axis of A_1 and A_2, and strike the target T at the center of the sphere S. A ring shield surrounds A_2 and is placed at a negative voltage to prevent stray electrons from striking the outside of the sphere. To simplify the voltage adjustments on the various electrodes, fixed values relative to F are applied to G_1, G_2, and A_1. A variable accelerating or retarding voltage is then applied between A_1 and A_2 to give the electrons the desired primary energy. This gun structure operates very satisfactorily over the whole voltage range used thus far (to 8000 volts). At the higher voltages of several thousand, a small change in the voltage of G_2 is required for maximum target current.

A galvanometer of sensitivity 6×10^{-10} ampere per mm measures the currents to T and S. A voltage between sphere and target to accelerate the secondary electrons leaving the target gives nearly the same results as are obtained with target and sphere at the same voltage, so the latter arrangement is used.

Observations have been made with the tube sealed to the vacuum system and also with a sealed-off tube. The vacuum system consists of a three-stage fractionating oil diffusion pump with a dry ice-acetone trap between pump and tube. The preliminary degassing consisted of two baking periods between which the trap was warmed to room temperature long enough to remove water vapor. The adequacy of vacuum was judged by the constancy of observations following separate heat treatment of the target, although an ionization gauge was available for reference. No observable change in δ

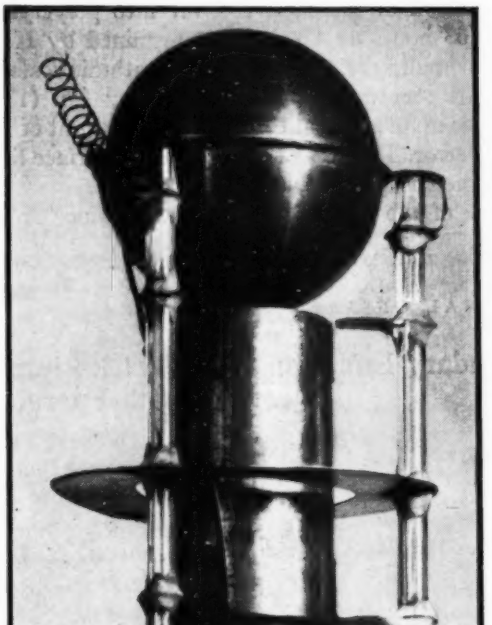


FIG. 2. Photograph of tube elements.

[3] H. E. Farnsworth, Rev. Sci. Instr. **21**, 102 (1950).

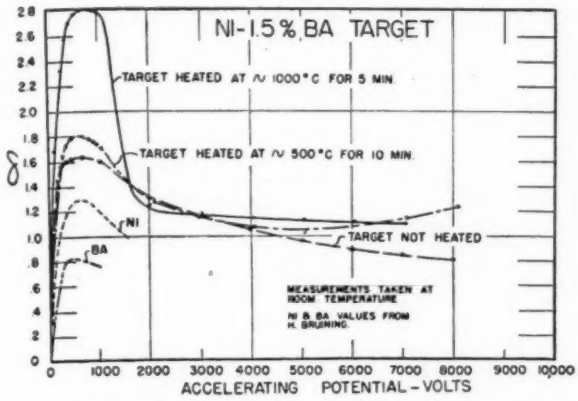

NI-1.5% BA TARGET
TARGET HEATED AT ∿ 1000°C FOR 5 MIN.
TARGET HEATED AT ∿ 500°C FOR 10 MIN.
NI
BA
TARGET NOT HEATED
MEASUREMENTS TAKEN AT ROOM TEMPERATURE
NI & BA VALUES FROM H. BRUINING.
δ
ACCELERATING POTENTIAL - VOLTS

FIG. 3. Secondary emission curves for cast alloy of nickel containing 1.5 percent barium.

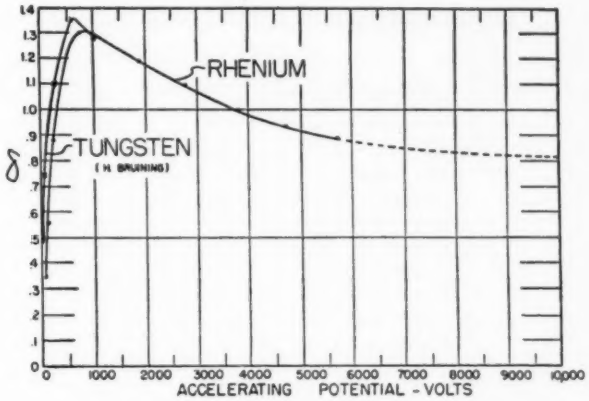


FIG. 4. Secondary emission curve for rhenium.

occurred over a period of several hours immediately following red heat treatment of target.

RESULTS AND DISCUSSION

Figure 3 shows a comparison of the results on secondary emission *vs* primary voltage for nickel-barium with those previously available for nickel and barium. After the usual preliminary baking of the tube but with no further heating of the target, the curve labeled "target not heated" was obtained. The target, in the form of a strip 3 mm wide, bent as in Fig. 1, was then heated by conduction for various times and at various temperatures up to bright red heat, the temperature being increased in steps as the time of heating was increased. Marked changes occurred in the height and shape of this curve. Several different targets cut from the same sample were investigated. In some cases the maximum in the emission curve decreased slightly with initial heating and then increased markedly with further heating at a higher temperature. In other cases this initial decrease was not observed but only the marked increase. The largest value of the emission maximum varied from sample to sample, the value of 2.8 shown in Fig. 3 being the largest one observed, but in every case the increase was very pronounced. The lowest observed value of the maximum is 1.64. It is also to be noted that the voltage at which the maximum occurs increases somewhat with heat treatment of target. Further, the marked increase is confined to the voltage range below 2000 volts, suggesting that this increase is due to a thin surface layer which has little effect on emission by higher energy electrons that generate secondaries from greater depths. It was also noted that when the target was in condition to yield the large emission below 2000 volts, the emission at the high voltages (4000–8000) was a function of the time of previous bombardment, the emission increasing with the bombardment. Additional heating of the target at still higher temperatures caused a decrease in the value of the emission maximum, so that the curve approached that obtained after only a relatively small amount of heating.

The maximum emission value of 2.8 for nickel-barium is much larger than that for either nickel or barium. While this may be due to the combination of the two metals, it should be noted that during the bulb sealing-in operation it was difficult to prevent completely the oxidation of the target; and, therefore, the increase in secondary emission ratio may be due to a thin surface layer of oxide. A value of the secondary emission ratio from barium-strontium oxide as high as 12 has been observed.[2] The subsequent decrease in emission probably results from a decrease in this surface layer of oxide.

Evaporation from the target is known to occur because of an increased secondary emission from the inner surface of the sphere. This is apparent when the sphere is negative with respect to the target.

Figure 4 shows the results for rhenium after degassing at red heat until an equilibrium curve was obtained. As is the case for other pure refractory metals, the effect of degassing is to lower the secondary emission curve. Some previous results in the low energy region for tungsten are shown for comparison.

The spherical collector S was constructed to permit energy distribution measurements by the retarding potential method. These measurements were obtained for both types of targets referred to above. For rhenium the results are similar to those for other refractory metals which have been reported in the literature. The most probable energy of low speed electrons from rhenium is approximately 5 ev. For the nickel-barium target the results are complicated by the fact that evaporation from the target increased the emission from the sphere, as mentioned above. This results in an error of energy distribution measurements by the retarding potential method. To decrease this error the inner surface of the sphere was coated with a layer of condensed smoke. While this reduces the emission from the sphere, it probably does not compensate entirely for the increased emission from the evaporated layer. For this reason the results on energy distribution of electrons from nickel-barium are not presented here.

Deterioration of Oxide-Coated Cathodes under Low Duty-Factor Operation*†

J. F. WAYMOUTH, JR.‡
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received July 27, 1950)

The behavior of oxide cathodes coated on a nickel alloy containing about 0.1 percent silicon has been investigated under 100 percent and zero duty-factor operation. The operation of these cathodes at zero duty factor was found to be highly favorable to the development of a high resistance in a layer of Ba_2SiO_4 located at the interface between the oxide and the core. The layer itself was present not only in cathodes that had aged without electron emission and had developed interface resistance, but also in cathodes that had aged with electron emission and had not developed interface resistance. A model is presented to account for the experimental observations. Evidence is also presented which shows that an "active" nickel alloy that does not lead to undue interface resistance formation may exist.

I. INTRODUCTION

OXIDE cathodes coated on an active-nickel-alloy core§ often fail before a thousand-hours life when they are operated at a low duty factor, due to the development of a high internal resistance. This failure has been recognized to be a major difficulty by the several groups now engaged in the construction of computers.[1] Tubes similar to those that failed in computer service have been found to last thousands of hours when they are operated with continuous emission current.

The internal cathode resistance of the failures could be accounted for by the hypothesis of an "interface resistance." The phenomenon of interface resistance has been investigated in connection with cathode sparking by Eisenstein,[2,3] Danforth and Goldwater,[4] Mutter,[5] Wright,[6] and others. These workers found that in cathodes coated on cores containing impurity concentrations of several percent, a layer of an insulating compound was formed at the interface between the oxide and the core, by chemical reaction between the oxide and an impurity.

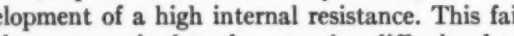


FIG. 1. Interface resistance measurement. a. No interface resistance. b. Interface resistance.

The dependence on duty factor, doubted by some, was a fundamental feature of the deterioration that could not be accounted for on the basis of existing knowledge of interface resistance. There was no adequate data to determine whether the deterioration was also associated with a loss of emission, as would be expected from the work of Wright.[6]

II. EXPERIMENTAL PROCEDURES

The research performed may be divided into two experiments. The first was a study of the behavior of cathodes coated on a commercial active-nickel alloy under high and low duty-factor operation; the second was an investigation into the dependence on the history of electron flow through the interface of the resistance and "activation energy" of an interface resistance on a high silicon nickel core.

A. Experiment I

Diodes similar in electrode structure to the ASTM "Standard Diode"|| were chosen for the study of cathode behavior under high and low duty-factor operation, since these diodes could be obtained in the necessary quantities from the Raytheon Manufacturing Company,¶ who produced them under carefully controlled conditions. With the exception of the alumina-coated tungsten heaters and the 699 nickel alloy cathode cores, all metallic parts of the diodes used were of 499 nickel alloy. Typical analyses of the two alloys are as follows:**

699 Si, 0.095 percent; Fe, 0.072; Mn, 0.085; Mg, 0.14; Cu, 0.006; Ti, 0.008

499 Si, 0.008 percent; Fe, 0.033; Mn, 0.008−; Mg, 0.009−; Cu, 0.009; Ti, 0.003.

Forty diodes were aged in four panels of ten tubes each at a cathode temperature of 1100°K. Diodes in

* Supported in part by the Signal Corps, the Air Force, and ONR.
† Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology.
‡ Now at Sylvania Electric Products, Inc., Salem, Massachusetts.
§ "Active" nickel alloys contain reducing agents in concentrations of the order of tenths of a percent. Such agents help activate cathodes coated on these alloys.
[1] Report R-139, Project Whirlwind Div. of M.I.T. Servomechanisms Laboratory.
[2] A. Eisenstein, J. Appl. Phys. **17**, 663 (1947).
[3] A. Eisenstein, J. Appl. Phys. **20**, 776 (1949).
[4] W. E. Danforth and D. L. Goldwater, J. Appl. Phys. **20**, 163 (1949).
[5] W. E. Mutter, Phys. Rev. **72**, A531 (1947).
[6] Wright, Proc. Roy. Soc. **A190**, 394 (1947).
|| Developed by American Society of Testing Materials Committee B-4, on Oxide Cathodes.
¶ Under their contract N7-onr-389.
** Furnished by Mr. James Cardell of Raytheon, private communication.

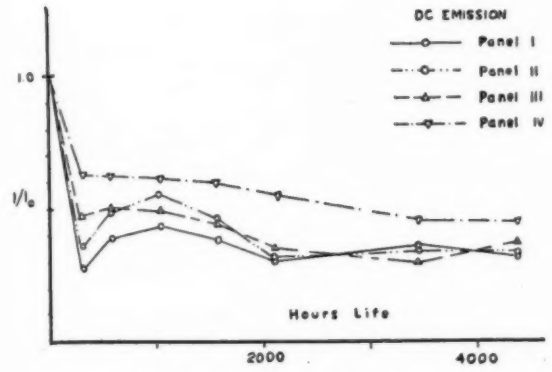


FIG. 2. Dc emission.

Panels I and II were aged with an electron emission of 60 ma per cm². Tubes in Panels III and IV were aged without electron emission. The tubes of Panel III had their anodes connected to their cathodes, while those of Panel IV had their anodes at −110 volts dc with respect to their cathodes, to simulate the effect of cut-off operation in a pentode or triode.

The tubes were tested periodically for dc emission, pulsed emission, and interface resistance. The dc emission measurement was made at a cathode temperature of about 600°K, with an applied anode potential of 4.5 volts. The average emission obtained at zero hours life was between thirty and forty microamperes.

The pulsed emission measurement was made at a cathode temperature of 940°K by applying to the diode a single one-millisecond 300-volt pulse from a thyratron pulser, and observing the emission by means of a synchroscope with a long-persistence screen. The average pulsed emission obtained at zero hours life was about 0.5 amp per cm².

The interface resistance measurement was made by pulsing the diode at constant current space-charge-limited in 2½ microsecond pulses and observing the voltage across the diode during the pulses with a synchroscope.†† Under these conditions, if the cathode has no interface resistance, the voltage pulse appearing across the diode is "square," and equal in height to the potential necessary to draw the constant pulse-current space-charge-limited, as shown in Fig. 1a. If the cathode has an interface resistance, a voltage builds up across this resistance during the pulse, as in Eq. (1):

$$V = iR[1-\exp(-t/RC)]. \quad (1)$$

R and C are the effective interface resistance and capacitance, respectively, and i is the constant pulse-current. Since the pulse length is 2½ microseconds and RC is of the order of a microsecond or less, V varies from zero to within a few percent of iR during the pulse.

When the diode is pulsed at constant current space-charge-limited, the potential difference between the surface of the cathode and the anode must remain constant during the pulse. Any change in the potential difference across the cathode must be compensated for by a change in the potential difference across the tube. Consequently, if the cathode has an interface resistance, the voltage pulse across the diode is no longer "square," but has the time-varying interface potential superimposed upon it, as shown in Fig. 1b.

To simplify the identification of an interface resistance, the measurement was made at a reduced cathode temperature (940°K). Because of the high negative temperature coefficient of interface resistance, the resistance at 940°K is a factor of five or more greater than at the normal operating temperature of 1050°K.

Note that a voltage pulse similar to that shown in Fig. 1b would be observed under the conditions of the interface resistance measurement if the cathode of the tube under test had no interface resistance, but its emission were decreasing with time during the pulse. To eliminate this uncertainty, the pulsed emission was measured at the same temperature as the interface resistance, and the variation of the emission with time was directly observed with the synchroscope. Since no tubes ever showed any detectable decrease in emission with time up to ten microseconds, voltage pulses having the shape shown in Fig. 1b could be used as evidence for the presence of interface resistance.

Results of Experiment I

The results of the dc and pulsed emission measurements are shown in Figs. 2 and 3, respectively. Each point represents the average emission from the tubes of the panel of ten, normalized to the average emission of the same tubes at zero hours life. Notice that, up to 4316 hours life, there is no significant difference in behavior among the four panels as far as emission is concerned.

The results of the interface resistance measurements, shown in Fig. 4 in the form of a survival curve, show a striking difference in behavior between tubes aged at zero duty factor and tubes aged at 100 percent duty-factor. Only one tube from Panels I and II failed, while all but three from Panels III and IV had failed at 4316

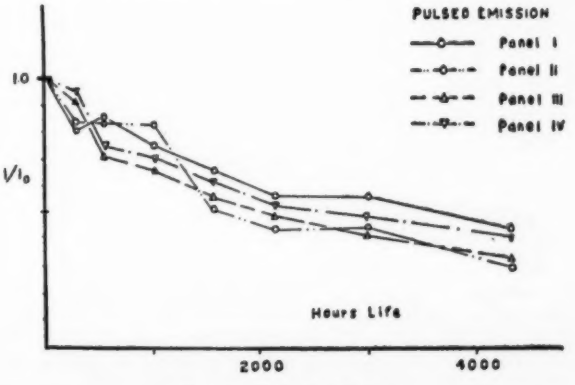


FIG. 3. Pulsed emission.

†† A brief description of the constant-current pulser is given in Appendix I.

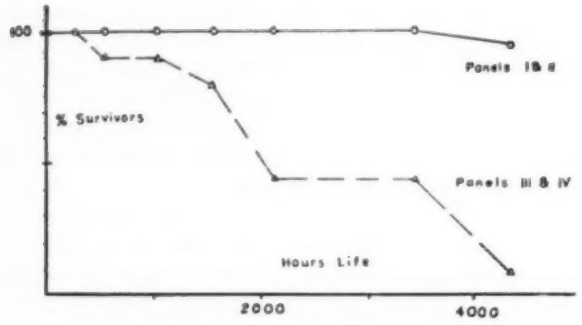


FIG. 4. Survival against interface resistance.

hours. The criterion for failure used in plotting Fig. 4 was the presence of any detectable interface resistance. The resistances of failures from Panels III and IV ranged from forty to about a thousand ohms. There was no difference between Panels III and IV.

At the conclusion of the aging program, x-ray diffraction and spectrochemical analyses were performed. For the identification of the interface compound by x-ray diffraction, samples were prepared by molding together (with Duco cement) scrapings from the interfaces of three cathodes whose coatings had been removed.‡‡ One sample was prepared from cathodes with interface resistance from tubes of Panels III and IV, and another from cathodes without interface resistance from tubes of Panels I and II. The layer of interface compound was visible in all of these cathodes as a grayish discoloration of the core; no difference between the two categories could be detected in the layer by visual examination.

Prints made from the patterns obtained from the two samples are shown in Fig. 5. They are quite similar; in each pattern, the heavy lines are caused by nickel, while the fine lines at low angle are due to Ba_2SiO_4. In each case, Ba_2SiO_4 is a major constituent of the scrapings from the interfaces. From this evidence, it may be concluded that a layer of Ba_2SiO_4 is present at the interface, not only in cathodes that have aged without electron emission and have developed interface resistance, but also in cathodes that have aged with electron emission and have not developed interface resistance.

Three categories were chosen for spectrochemical analysis:

A. Tubes without interface resistance (Panels I and II);
B. Tubes with interface resistance (Panels III and IV);
C. Tubes without interface resistance (Panels III and IV).

Four samples were obtained from each of these categories, as follows:

a. Surface of anode;
b. Bulk of anode;
c. Cathode coating plus surface of nickel core;
d. Bulk of nickel core.

‡‡ The identification of the interface compound was attempted from patterns taken using the entire core (with coating removed) as a sample. However, the layer of interface compound was too thin to give detectable lines on the patterns.

TABLE I. Analysis of anodes.

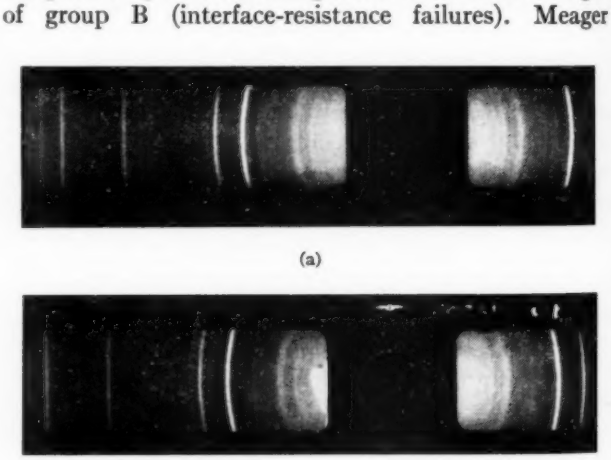
Group	Sample	Impurity concentration in % Si	Fe	Mn	Mg	Cu	Ti
A	(a)	0.029	0.027	0.062	0.011	0.0065	nil
	(b)	0.012	0.010–	nil	0.002–	0.0042	nil
B	(a)	0.026	0.028	0.12	0.038	0.009	nil
	(b)	0.012	0.010–	nil	0.002	0.0037	nil
C	(a)	0.036	0.022	0.057	0.042	0.0052	nil
	(b)	0.012	0.010–	nil	0.002–	0.0046	nil
#499 alloy[a]		0.008	0.033	0.008	0.009–	0.009	0.003–

Sample (a) anode surface.
Sample (b) anode volume.

[a] Typical analysis, furnished by Mr. James Cardell of the Raytheon Manufacturing Company, private communication.

The results of the spectrochemical analyses are shown in Tables I and II. Note that in every case, the concentration of impurities is greater in the anode surface than in the bulk of the anode. This may be ascribed to a contamination of the anode surface by impurities evaporated from the cathode. A comparison of analyses for group A and group B indicates that it is unlikely that the results obtained under the interface-resistance measurement are due to a systematic difference in impurity content between tubes of Panels I and II and tubes of Panels III and IV.

A significant difference in impurity content is shown in the sample labelled "cathode coating plus surface of nickel core" of group C, which consisted of tubes that aged without electron emission without developing interface resistance. The iron concentration in this sample is only one-tenth as great as in the same sample of group B (interface-resistance failures). Meager



(a)

(b)

FIG. 5. X-ray diffraction patterns of interface compound. a. Cathodes with interface resistance. b. Cathodes without interface resistance.

TABLE II. Analysis of cathodes.

Group	Sample	Impurity concentration in % Si	Fe	Mn	Mg	Cu	Ti
A	(c)	0.38	0.16	0.048	0.043	0.0084	0.027
	(d)	0.048	0.10	0.054	0.011	0.007	0.010–
B	(c)	0.17	0.10	0.035	0.057	0.0056	0.050
	(d)	0.042	0.084	0.044	0.0084	0.0074	0.010
C	(c)	0.27	0.010	0.052	0.059	0.0046	0.022
	(d)	0.032	0.096	0.070	0.011	0.0074	0.010–
#699 alloy[a]		0.095	0.072	0.085	0.14	0.006	0.008

Sample (c) coating plus surface of nickel core.
Sample (d) bulk of nickel core.

[a] Typical analysis, furnished by Mr. James Cardell of the Raytheon Manufacturing Company, private communication.

though this evidence is, it seems to indicate that an alloy containing as much silicon and manganese as 699, but far less iron, might be significantly better as a base-metal for cathodes to be operated under low duty-factor conditions.

B. Experiment II

The purpose of the second experiment was to investigate the dependence on the history of electron flow of the resistance and "activation energy" of an interface resistance on a high silicon nickel core. A "sandwich" diode was constructed, in which a wafer of (BaSr)O was compressed between two flat electrodes, as shown in Fig. 6. The electrode labelled "K_1" was of silicon-nickel alloy (about 5 percent Si), while "K_2" was of "Grade A" nickel (less than 0.3 percent Si). Two 0.7-mil platinum wire probes imbedded in the wafer, as shown in Fig. 6, permitted the determination of the potential difference across the interface by the "double-probe" method which has been used by Eisenstein[2] and others.

The wafer was constructed by spraying a coating on K_1 in three layers with probes between the layers; a single-layer coating was sprayed on K_2, and the two were mounted in contact to form the wafer. Since there was apparently a certain amount of potential difference across the junction of the coatings, the probes were used only to measure the potential across the K_1 interface. The temperature of this interface was measured with a tungsten-nickel thermocouple. One leg of this thermocouple was the K_1 electrode itself, while the other was a

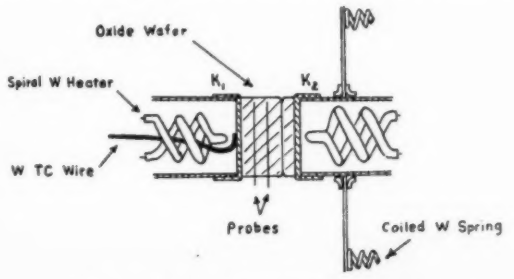


FIG. 6. Schematic of sandwich diode.

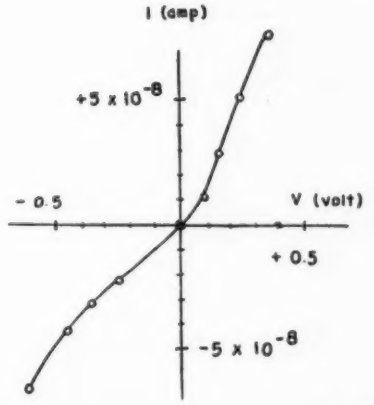


FIG. 7. Interface $I-V$ characteristic (790°K).

three-mil tungsten wire welded to the back of the electrode, as shown in Fig. 6.

In the process of conversion of the carbonates, great care was taken to raise the temperature of the wafer slowly with the pressure monitored by an ion gauge, to prevent excessively rapid liberation of CO_2. This was felt to be necessary because of the relatively large volume of the wafer, and the small surface area it presented to the vacuum. After the conversion of the carbonates, the wafer was activated by the passage of electron current at a temperature in excess of 1000°K until no further reduction in resistance took place. At this time, the resistance at 900°K of the wafer (0.2 cm² area of cross section, 0.5 mm thick) was about a megohm. The pressure at seal-off was less than 5×10^{-8} mm.

A typical current-voltage characteristic obtained for the K_1 interface is shown in Fig. 7. Note the rectifying nature of the interface resistance, similar to that observed by Mutter.[5] The high resistance direction is the direction of electron flow from the core to the oxide. To obtain from these characteristics a value of the "activation energy," use was made of the fact that in the theory of rectifying contacts, the current can be expressed as a function of the voltage by an expression of the form§§

$$i=A(T)f(V/T)\exp(-\epsilon\Phi/kT), \qquad (2)$$

where Φ is the activation energy in electron-volts, and $A(T)$ varies as T^n, where n is of the order of one or two; k is the Boltzmann constant.

Measurements were taken of the current-voltage characteristic in the high resistance direction at three different temperatures; as few points as possible were read on each curve to avoid difficulties with lack of reproducibility. The logarithm of the current observed at a fixed value of V/T for each of the temperatures was plotted as a function of $1/T$. From the slope of the best straight line that could be drawn through the three points, the activation energy was obtained.

The result of a typical measurement of this kind is shown in Fig. 8. The numbered points are the repro-

§§ See, e.g., Torrey and Whitmer, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1948), p. 77.

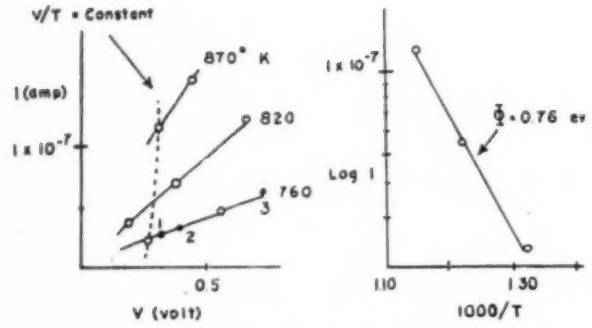


FIG. 8. Measurement of activation energy.

ducibility check points, which were taken with the same potential applied across the wafer after the measurements at 760°K, 820°K, and 870°K, respectively. Note that although the oxide has tended to activate during the measurements, the state of the interface has not changed.

The tube was aged at a wafer temperature of 1000°K as follows:

0–503 hours; electrons flow from oxide to K_1 interface;
503–668 hours; electrons flow from K_1 interface to oxide;
668– hours; electrons flow from oxide to K_1 interface.

The currents that flowed during the ageing were about ten or twenty microamperes, and the potential difference across the interface was about one or two volts. If a value of 10^{-3} cm is assumed for the thickness of the interface region, the electric field in this region can be calculated to have been of the order of one or two thousand volts per cm.

The variation of the activation energy of the interface resistance with time during the aging is shown in Fig. 9. As long as electrons flowed from the oxide to the interface, the activation energy increased, but when electrons flowed from the interface to the oxide, the activation energy decreased sharply. A similar behavior was shown by the resistance, as may be seen in Fig. 10.

III. INTERPRETATION

The results that have been obtained have posed a particularly serious question: how can a layer of Ba_2SiO_4 (an insulator) at the interface between the oxide and the core have a *high resistance* after ageing with no current flow, and have a *low resistance* after ageing with electron flow from core to oxide? To answer this question, which I believe to be fundamental to the understanding of the behavior of oxide cathodes under low duty-factor operation, I will propose a model of the interface region and show that it is capable of accounting qualitatively for the observations of both Experiments I and II.

During the process of activation of a cathode coated on a silicon-bearing nickel core, a layer of Ba_2SiO_4 forms at the interface by a reaction between silicon in the core and BaO of the coating. This reaction is probably at least partly responsible for the formation of the stoichiometric excess of barium necessary for an active cathode. Some of this excess of barium undoubtedly remains in the layer of interface compound, providing it with donor centers which transform it from an insulator to an "excess" or N-type semiconductor. The conductivity of the Ba_2SiO_4 in this state is high enough that the cathode has a negligible interface resistance.

The energy level configuration of the interface region in this state would be as shown in Fig. 11. The diagram has been drawn in accordance with the views of Slater[7] concerning the junction between conductors and semiconductors, with the *a priori* assumption that the work-function of the activated Ba_2SiO_4 does not differ greatly from the work-function of the activated oxide, but that the work-function of the nickel is greater than either. In all probability, the Schottky "depletion-layer" barrier at the interface between the nickel and the Ba_2SiO_4 has a negligible resistance, because the boundary is undoubtedly not abrupt, as shown, but much more "diffuse."

As the cathode is aged, the layer of Ba_2SiO_4 thickens to the point where reaction between silicon of the core and BaO of the coating becomes negligible. The behavior of the concentration of donor centers in the layer during the aging will depend on whether an electric field is present, because the donor centers behave on the average as positive charges.

If no electron emission is drawn from the cathode, there is no electric field in the interface region, and the donor centers can migrate out of the layer of interface compound by thermal diffusion. The concentration of donor centers in the layer is reduced to the extent that it is insufficient to maintain a negligible resistance; the cathode has then developed an interface resistance.

If the cathode is aged delivering electron emission, an electric field directed from oxide to core exists in the interface region. Field-induced migration of donor centers into the layer of interface compound and thermal diffusion of donor centers out of the layer of

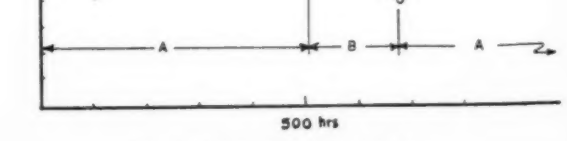


FIG. 9. Activation energy *vs* time.

[7] J. C. Slater, Phys. Rev. **76**, 1592 (1949).

interface compound are then in competition. The equilibrium concentration of donor centers in the Ba_2SiO_4 layer is greater than it would be if thermal diffusion were the only migration process. If the field is strong enough, the equilibrium concentration of donor centers is sufficient to maintain the conductivity of the interface compound. The cathode does not develop an interface resistance.

The energy level configuration of the interface region of a cathode that aged delivering electron emission and did not develop interface resistance would probably be quite similar to that shown in Fig. 11, with the exception that the layer of interface compound would be thicker. For a cathode that did develop interface resistance, the energy level configuration would be as shown in Fig. 12.

In Fig. 12, the levels of the Ba_2SiO_4 have been shifted upward relative to the levels of the nickel and the oxide, because of the dependence of the Fermi level (μ) in the Ba_2SiO_4 on the concentration of donor centers. Since the resistance of the Schottky "depletion-layer" barrier can still be neglected, the activation energy of the interface resistance would be equal to Φ, the energy difference in electron volts between the Fermi level (μ) and the bottom of the conduction band in the Ba_2SiO_4, as shown in the figure.

If the cathode were aged with electrons flowing from the oxide to the core, an electric field directed from the core to the oxide would exist in the interface region. The field-induced migration of donor centers would then be in such a direction as to reduce the concentration in the interface compound, and would aid rather than retard the development of an interface resistance. Furthermore, a comparison of Figs. 11 and 12 shows that as the concentration of donor centers decreases and the resistance increases, the activation energy also increases.

This is exactly what was observed in Experiment II; as long as the interface was aged with electrons flowing from the oxide to the core, the resistance and activation energy increased. The fact that resistance and activation energy could be reduced by the passage of electron current from the core to the oxide is also in agreement with the features of this model, since there is no reason why a low concentration of donor centers could not be

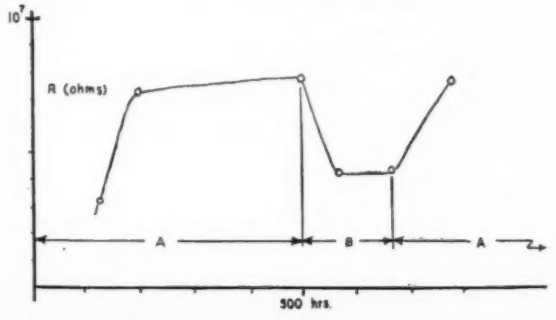


FIG. 10. Resistance *vs* time (790°K).

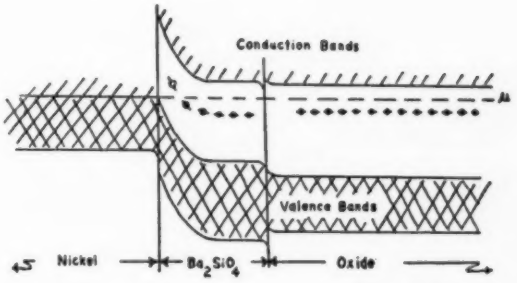


FIG. 11. Energy level system early in life.

increased by the establishment of field-induced migration in the proper direction.

To account for one of the results of Experiment II, it is necessary to modify the model slightly, without altering the fundamental features already outlined. An examination of Figs. 9 and 10 shows that the resistance is not related to the activation energy by the simple expression

$$R\alpha \exp(\epsilon\Phi/kT), \tag{3}$$

as would be expected on the basis of Eq. (2). The variation observed in R is not nearly as great as would be expected from the variation observed in Φ.

This can be accounted for by the fact that the interface layer is probably not at all homogeneous, but varies in thickness and in donor center concentration. Conduction through it depends on parallel conduction through many "patches" of different thickness, conductivity, and activation energy. The experimental measurements of resistance and activation energy represent two different methods of averaging over the patches. Since the degree of inhomogeneity, on which the resistance strongly depends, can be quite different for two states of the interface barrier having the same observed activation energy, it is not surprising that the resistance and activation energy fail to be related in the simple manner expressed by Eq. (3).

IV. SUMMARY AND CONCLUSIONS

The results obtained in this work have shown that the anomalous deterioration that occurs in oxide cathodes coated on silicon-bearing nickel cores when they are operated at a low duty factor is due entirely to the development of a high resistance in a layer of Ba_2SiO_4 located at the interface between the oxide and the core. The layer itself was found to be present not only in cathodes that had developed interface resistance, but also in cathodes that had not. The development of a resistance in this layer was heavily favored by the absence of electron emission from the cathode during life.

Further investigation of the properties of an interface resistance on a high silicon nickel core has shown the resistance and activation energy to be strikingly dependent on the history of electron flow through the interface. A model has been presented that appears

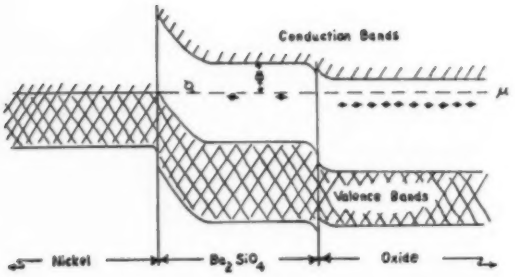


FIG. 12. Energy level system of cathode with interface resistance.

capable of accounting for all the experimental observations, and of providing a framework for the understanding of the fundamental features of the deterioration that has been under study.

The application of this understanding to the engineering problem of developing cathodes that do not deteriorate when they are operated at a low duty factor leads to two main avenues of approach.

(1) Use nickel containing less silicon and other impurities; the layer of interface compound will be thinner, and will have less resistance.

(2) Search for other substances that can be added to the interface compound to provide it with permanent donor centers that will persist in the layer irrespective of duty factor.

The first of these avenues is currently being exploited by most of the major tube manufacturers, although it suffers from two faults. First, there is enough variation in impurity content among different melts of what is nominally the same alloy that different production lots of the same tube can have widely different life expectancies. Second, the use of alloys containing appreciable amounts of impurities (i.e., "active" alloys) is often highly desirable as an aid to easy activation of the cathodes.

That the second avenue is not beyond the realm of possibility appears to be indicated by the results of the spectrochemical analyses of diodes that had aged 4316 hours with no electron emission without developing interface resistance. The cathodes of these diodes were found to contain far less iron in the "cathode-coating-plus-surface-of-nickel-core" sample, which contains the interface, than cathodes that had developed interface resistance. It is conceivable that in cathodes containing very little iron, the presence of $BaMnO_4$ in the Ba_2SiO_4 lattice might provide just such permanent donor centers. The dependence on the iron content stems from the fact that the coexistence of $BaMnO_4$ and Fe side by side at a temperature of 1100°K for 4000 hours would be extremely improbable, because the MnO_4^{2-} ion is a strong oxidizing agent.

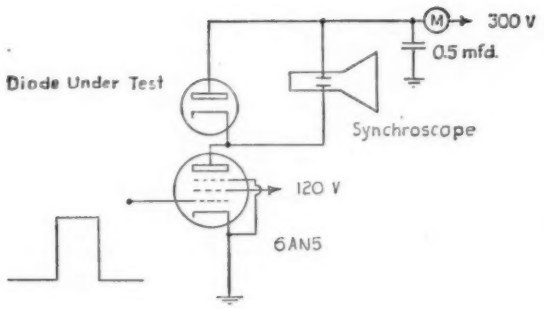


FIG. 13.

However meager the data, and whatever the mechanism may be, these results appear to indicate at first sight that a nickel alloy containing all impurities except iron to the extent that they are found in 699 nickel alloy might be a satisfactory active nickel alloy for cores of cathodes to be operated at a low duty factor.

V. ACKNOWLEDGMENTS

The author is deeply indebted to Mr. James Cardell, of the Raytheon Manufacturing Company, who performed the spectrochemical analyses, and provided the diodes used in Experiment I; to Mr. H. B. Frost of the M.I.T. Servomechanisms Laboratory, who assisted greatly with the pulsed emission and interface resistance measurements of Experiment I; to Mr. C. B. Walker of the M.I.T. X-Ray Laboratory, who assisted with the x-ray diffraction analyses; to the technicians and glass-blowers of the Research Laboratory of Electronics; to his associates in the M.I.T. Physical Electronics Group for many helpful discussions; to Prof. W. B. Nottingham of M.I.T., who supervised this research, for many helpful criticisms and much helpful advice.

APPENDIX I. CONSTANT-CURRENT PULSER

A diagram of the essential features of the constant current pulser is shown in Fig. 13. The grid of the 6AN5, which is normally biased beyond cutoff, is supplied with flat-topped pulses from a voltage-pulse generator. The pulse amplitude is adjusted until the current through the diode during the pulse is 50 ma. Under these conditions, the potential across the diode during the pulse is about 45 volts.

The plate current of a 6AN5 drawing 50 ma with 120 volts screen voltage and 250 volts plate voltage is approximately independent of plate voltage. When the total drop across the diode increases by four or five volts during the pulse because of interface resistance, the plate voltage of the 6AN5 is reduced by the same amount, but the plate current is not affected. The current through the diode is constant during the pulse irrespective of whether or not the diode has an interface resistance.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Associated Legendre Polynomial Approximations

R. LANDAUER*
Harvard University, Cambridge, Massachusetts
(Received July 3, 1950)

Approximations for the associated Legendre Polynomials are derived by a phase integral method. The method is an extension of the WBK method, applicable to separable multidimensional wave propagation problems.

INTRODUCTION

IT has frequently[1,2] been pointed out that $Ae^{iS/\hbar}$ is an approximate solution of the Schroedinger equation. Here S is the solution of the corresponding classical Hamilton-Jacobi equation. A^2 is the density of an assembly of classical particles moving along the paths associated with S. In one dimension this approximation reduces to the WBK approximation. The many dimensional approximation was used by Gordon[3] in discussing Coulomb scattering. We shall use it here to derive an approximation for the associated Legendre Polynomials, $P_l^m(\cos\theta)$. Since these polynomials are elementary functions, the need for approximations is not as great as with other harmonics. It will, nevertheless, be useful to have an expression by which all $P_l^m(\cos\theta)$ can be approximated, if the number of steps involved in calculating the approximation is independent of l and m.

A well-known approximation for $P_l^m(\cos\theta)$, if $l\gg m$, and if θ is not too near 0 or π, is[4]

$$P_l^m(\cos\theta)=(-l)^m\left(\frac{2}{l\pi\sin\theta}\right)^{\frac{1}{2}}\times\sin\left(\left[l+\tfrac{1}{2}\right]\theta+m\frac{\pi}{2}+\frac{\pi}{4}\right). \quad (1)$$

The more general case, where m can be comparable with l, was treated by Nicholson.[5] His approximations do not have the proper number of zeros and are therefore unsatisfactory.

Theoretically, an approximation to the associated Legendre polynomials can be obtained by applying the ordinary WBK method directly to the differential equation for the Legendre polynomials. The resulting phase integrals become very complicated, and in addition have to be modified in a manner similar to that used in treating the radial part of the wave equation by the WBK method.[6]

THE METHOD

The time independent form of the Hamilton-Jacobi equation can be separated in spherical coordinates if the potential, V, is a function of r only. The solution has the form

$$S=R(r)+\Theta(\theta)+\gamma\phi. \quad (2)$$

In order that $Ae^{iS/\hbar}$ be a single valued solution, γ must be an integral multiple of $\hbar$. The Hamilton-Jacobi equation then separates into

$$d\Theta/d\theta=\pm(\beta^2-m^2\hbar^2/\sin^2\theta)^{\frac{1}{2}}, \quad (3)$$

$$dR/dr=\pm(2m(E-V)-\beta^2/r^2)^{\frac{1}{2}}, \quad (4)$$

β is a separation constant; $m\hbar=\gamma$. We will here be concerned primarily with Eq. (3). It has two solutions, corresponding to the two sign possibilities. For $\theta_0<\theta<\pi-\theta_0$, where $\sin\theta_0=m\hbar/\beta$, $d\Theta/d\theta$ is real. For $\theta<\theta_0$ and $\theta>\pi-\theta_0$, $d\Theta/d\theta$ is imaginary. $\theta=\theta_0$ and $\theta=\pi-\theta_0$ are then turning points at which the classical motion must stop.

Equation (3) can be integrated by the methods given in Appendix II of Born's *Vorlesuugen über Atommechanik*.[7] For the phase in the allowed region of motion this gives

$$\int_{\theta_0}^{\theta}(\beta^2-m^2\hbar^2/\sin^2\theta)^{\frac{1}{2}}d\theta=\vartheta(\theta). \quad (5)$$

$$\vartheta(\theta)=\frac{\beta}{\hbar}\cos^{-1}\frac{\mu}{(1-k^2)^{\frac{1}{2}}}-m\tan^{-1}(1-k^2-\mu^2)^{\frac{1}{2}}/k\mu, \quad (6)$$

where $\mu=\cos\theta$ and $k=\sin\theta_0$. Similarly we have, for $\theta<\theta_0$

$$\int_{\theta}^{\theta_0}\left|\frac{d\Theta}{d\theta}\right|d\theta=d(\theta). \quad (7)$$

$$d(\theta)=m\tanh^{-1}(\mu^2+k^2-1)^{\frac{1}{2}}/k\mu-\frac{\beta}{\hbar}\cosh^{-1}\mu/(1-k^2)^{\frac{1}{2}}. \quad (8)$$

We have now found the phase and must also evaluate the amplitude. The amplitude can be written

$$A(r,\theta,\phi)=A_r(r)A_\theta(\theta)A_\phi(\phi). \quad (9)$$

The symmetry of the problem requires that A_ϕ be constant. Consider the surfaces of constant θ. Since we

* Present address: Lewis Flight Propulsion Laboratory, NACA, Cleveland, Ohio.

[1] J. H. Van Vleck, Proc. Nat. Acad. Sci. **14**, 178 (1928).
[2] J. Frenkel, *Wave Mechanics, Advanced General Theory* (Oxford University Press, London, 1934).
[3] W. Gordon, Z. Physik **148**, 180 (1928).
[4] Jahnke and Emde, *Tables of Functions* (Dover Publications, New York, 1945), p. 117.
[5] J. W. Nicholson, Quart. J. Math. **41**, 241 (1910).
[6] E. C. Kemble, *Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1937), p. 107.
[7] M. Born, *Vorlesuugen über Atommechanik* (Verlag. Julius Springer, Berlin, 1925).

are dealing with a stationary process the flow of particles across all these surfaces must be the same. The density of the particles is proportional to $A_\theta^2 A_r^2$. The component of the velocity normal to a surface of constant θ is $1/mr(d\Theta/d\theta)$, where m is the mass of one of the particles. The flow across an element of area $r\sin\theta d\phi dr$ is then

$$A_r^2 A_\theta^2 \frac{1}{m}\frac{d\Theta}{d\theta}\sin\theta d\phi dr. \tag{10}$$

The net flow across the surface of constant θ is then

$$\frac{A_\theta^2}{m}\frac{d\Theta}{d\theta}\sin\theta\int\int d\phi dr A_r^2. \tag{11}$$

In order that this be independent of θ we must have

$$A_\theta(\theta)=\frac{1}{(\sin\theta)^{\frac{1}{2}}}\frac{1}{(d\Theta/d\theta)^{\frac{1}{2}}}. \tag{12}$$

In the classically allowed region, $\theta_0<\theta<\pi-\theta_0$, there are therefore two approximate solutions for the part of the wave function depending on θ. They are

$$\frac{1}{(\sin\theta)^{\frac{1}{2}}}\frac{1}{[\vartheta'(\theta)]^{\frac{1}{2}}}e^{\pm i\vartheta(\theta)}. \tag{13}$$

In the region $\theta<\theta_0$ there are also two solutions

$$\frac{1}{(\sin\theta)^{\frac{1}{2}}}\frac{1}{|d'(\theta)|^{\frac{1}{2}}}e^{\pm d(\theta)}. \tag{14}$$

This leaves us with a connection problem. If we ignore the $1/(\sin\theta)^{\frac{1}{2}}$ factor, (13) and (14) have the form of a one-dimensional WBK approximation and are in fact the approximations for $y(\theta)$, where $y(\theta)$ satisfies

$$\frac{d^2y}{d\theta^2}-\frac{m^2}{\sin^2\theta}y+\frac{\beta^2}{\hbar^2}y=0. \tag{15}$$

Since $1/(\sin\theta)^{\frac{1}{2}}$ does not have a branch point at $\theta=\theta_0$, and is single valued, it will suffice to connect the remaining factors by the usual connection formulas.[8]

Of the two solutions given by (14) only

$$\frac{1}{(\sin\theta)^{\frac{1}{2}}}\frac{1}{|d'(\theta)|^{\frac{1}{2}}}e^{-d(\theta)}, \tag{16}$$

remains finite at $\theta=0$, and we must connect to it at $\theta=\theta_0$. Similarly at $\theta=\pi-\theta_0$. This requires that

$$\frac{1}{\hbar}\int_{\theta_0}^{\pi-\theta_0} p_\theta d\theta=\vartheta(\pi-\theta_0)-\vartheta(\theta)=(n+\tfrac{1}{2})\pi. \tag{17}$$

This is satisfied whenever

$$\beta=(l+\tfrac{1}{2})\hbar, \tag{18}$$

where l is an integer greater than m.

Our approximate solution for $P_l^m(\cos\theta)$ in the allowed region is therefore

$$\frac{C}{(\sin^2\theta-\sin^2\theta_0)^{\frac{1}{4}}}\cos\left(\vartheta(\theta)-\frac{\pi}{4}\right). \tag{19}$$

For $\theta<\theta_0$ the approximate solution is

$$\frac{1}{2}\frac{C}{(\sin^2\theta_0-\sin^2\theta)^{\frac{1}{4}}}e^{-d(\theta)}; \tag{20}$$

while for $\theta>\pi-\theta_0$ the solution becomes

$$(-1)^{l-m}\frac{1}{2}\frac{C}{(\sin^2\theta_0-\sin^2\theta)^{\frac{1}{4}}}. \tag{21}$$

These formulas still contain an unevaluated normalization constant. Apart from this, the approximations have the general shape of $P_l^m(\cos\theta)$. At $\theta=0$ the approximation goes to zero as θ^m, which is correct. The other solution given by (14), which we did not use, varies as θ^{-m} near $\theta=0$, which is characteristic of $Q_l^m(\cos\theta)$. Furthermore the approximation has the proper number of zeros, which is $l-m$, not counting the ones at $\theta=0$ and $\theta=\pi$. The most prominent difference between the various associated Legendre polynomials having the same value of $l-m$, is that those with higher values of m have their oscillations more closely packed about $\theta=\pi/2$. This feature is also given by our approximation. In the case $l\gg m$ our approximation reduces to (1) in the region where (1) is applicable.

If $P_l^m(\cos\theta)$ is defined as

$$P_l^m(\cos\theta)=\sin^m\theta\frac{d^mP_l(\cos\theta)}{d(\cos\theta)^m}, \tag{22}$$

then $P_l^m(\cos\theta)$ is positive for θ near 0. This will be satisfied by our approximation if C is chosen real and positive.

The numerical value of C can be evaluated by the method discussed by Furry.[9] The Legendre polynomials obey

$$\int_0^\pi \sin\theta[P_l^m(\cos\theta)]^2 d\theta=\frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}, \tag{23}$$

which gives us

$$\int_{\theta_0}^{\pi-\theta_0}\frac{A_\theta^2}{2}\sin\theta d\theta=\frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}, \tag{24}$$

[8] See reference 6, p. 94.

[9] W. H. Furry, Phys. Rev. **71**, 360 (1947).

and therefore

$$C=\left[\frac{4}{\pi(2l+1)}\frac{(l+m)!}{(l-m)!}\right]^{\frac{1}{4}}. \quad (25)$$

We can expect our approximation to be accurate in the limit of closely spaced nodes. They will be closely spaced if $l-m$ is large. But as in any phase integral approximation, a reasonable representation of the function is obtained even for low quantum numbers, or small $l-m$.

If the value of $\beta=(l+\frac{1}{2})\hbar$ is now used in (4), the radial factor can be treated by a method similar to the one we have used above. The resulting approximation will be identical with that obtained by the usual method of applying the WBK approximation to radial motion.

A more formal and rigorous derivation of the approximations can be given along lines similar to those used here.

The author wishes to thank Professor Furry for his help and advice and the AEC for a Predoctoral Fellowship.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

A Pulse Method of Determining the Energy Distribution of Secondary Electrons from Insulators

KENNETH G. MCKAY
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received August 2, 1950)

A novel method is described of determining the energy distribution of emitted secondaries from an insulator. This is based on an analysis of the transient resulting from pulse bombardment. The analysis is simplest when leakage through the target is negligible, but the effect of leakage is also treated. Space charge limitation of the emitted current is assumed to be negligible.

1. INTRODUCTION

RECENT studies of secondary emission by pulse methods have demonstrated the power of this approach to the problem of determining the secondary emission yield from insulators.[1-3] Essentially, the time interval between pulses of the primary bombarding current is utilized to restore the surface potential of the target to a known value before the incidence of the succeeding pulse. Thus, the effects of surface charging can be accounted for in a known manner. The published measurements using this technique have been concerned with the determination of only the secondary emission yield. The purpose of this paper is to show that a transient analysis of the pulse can give other information about the target and, in particular, can give a rough measure of the energy distribution of the emitted secondaries.

Figure 1 is a schematic of the usual circuit for such pulse measurements. A pulse of electrons from the electron gun strikes the target, and the secondaries are picked up by the collector. The metallic contact on the back of the target is connected through an output resistor R_0 to ground, and the voltage developed across R_0 is amplified and displayed on an oscilloscope. At the beginning of the pulse the bombarded face of the target is at ground potential. If the yield is greater than unity, the target begins to charge positively. Provided the leakage through the target is high, the target surface will reach an equilibrium potential which is less than that of the collector; this is the usual situation in such measurements. However, if the leakage is negligible, the target will charge up to a potential greater than that of the collector. The secondaries are then in a retarding field and the slowest will be turned back to the target. The target will continue to charge up until only the fastest secondaries can reach the collector and the effective yield has dropped to unity. During this latter phase, the rate of change of target potential, which depends on the current of secondary electrons, is therefore a function of the energy distribution of the secondaries. Moreover, the target surface is coupled electrically to the amplifier input through the impedance of the crystal. Thus, the problem can be expressed as

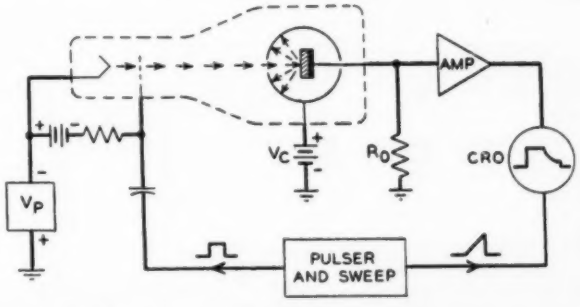


FIG. 1. Schematic of typical circuit for pulse measurements of secondary emission.

[1] H. Salow, Z. techn. Phys. **21**, 8 (1940).
[2] J. B. Johnson, Phys. Rev. **66**, 352 (1944); **73**, 1058 (1948).
[3] M. A. Pomerantz, J. Franklin Inst. **241**, 415 (1946).

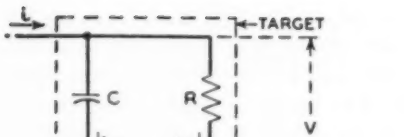


FIG. 2. Basic circuit of target and load resistance.

follows: given the observed transient as a function of time, we wish to determine the energy distribution of the secondaries and other relevant quantities such as the yield, the resistivity, and the dielectric constant of the target. Essentially, this is a retarding field measurement of the energy distribution in which independent control of the field is not available.

2. ASSUMPTIONS AND SYMBOLS

To facilitate the analysis we make a number of assumptions which appear reasonable from the experimental point of view:

(1) The collector is spherical and the target is small, so that the secondaries are in an essentially central field, and we can neglect the capacity between the target surface and the collector.

(2) We can neglect loss of secondaries through the collector entrance hole and can neglect tertiaries produced at the collector.

(3) The target is uniform and plane parallel with negligible edge effects. This assumption is reasonable if the target thickness is much less than the diameter of the bombarded area and if the bombarding beam covers essentially all of the exposed surface of the target.

(4) The bombarding voltage is large compared with the collector voltage, so that the primaries always strike the target with essentially the same energy.

(5) The input circuit and amplifier do not cause an appreciable modification of the transient by their band width.

(6) We can neglect space charge effects in the collector space. This condition can be attained with sufficiently small currents together with an amplifier with a good signal-to-noise ratio as discussed in Sec. 5.

(7) We assume that the target surface is at ground potential immediately before the onset of the primary pulse. Various methods have been used to restore the surface potential between pulses. Johnson[2] used a long interval between pulses and relied upon the small but finite conductivity of the target itself to restore the surface potential. Salow[1] flooded the target with an intense dc electron beam from a second gun at such a potential that the dc bombarding energy was less than that of the first cross-over point, and thus the target surface remained essentially at the potential of the cathode of the dc emitter. One could substitute a second pulsed beam for the dc beam if desired. No matter what the properties of the target may be, it is always possible, by one means or another, to restore the surface to a known potential before the onset of the pulse under observation. For these calculations we shall assume that by some means or other this has been done.

(8) We shall treat the bombarding current as a step function such that for $t<0$, the bombarding current is zero, and for $t>0$, the bombarding current is a constant.

The following symbols are used:

V_p = energy of the primary electrons in electron volts.
V_c = collector potential with respect to ground in volts.
V = potential of the bombarded target area relative to the metal base in volts.
R_0 = output resistance between target contact and ground in ohms.
E = voltage developed across R_0, which feeds the amplifier, in volts.
ρ = specific resistivity of target in ohm cms.
κ = dielectric constant of target.
l = thickness of target in cms.
A = bombarded area of target in cms.^{2}
$C=0.0885\kappa A/l$ is the capacity between front and back surfaces of the target in $\mu\mu$f.
$R=\rho l/A$ is the internal resistance between front and back surfaces of the target in ohms.
i_p, i_s = the primary and secondary currents respectively (amp).
$j_p, j_s = i_p/A,\ i_s/A$ = the primary and secondary current densities respectively immediately in front of the target in amp/cm^2.
$\delta = i_s/i_p$ = the secondary yield.
E_s = the energy of emission of secondary electrons in electron volts.
$P(E_s)$ = the probability distribution of the energy of emission of all secondary electrons per incident primary; i.e., $P(E_s)dE_s$ is the probability that a secondary will be emitted with energy between E_s and E_s+dE_s as the result of the incidence of one primary electron.

3. SECONDARY EMISSION PULSE RESPONSE

The basic problem to be solved is the behavior of a series parallel RC circuit fed by a current as in Fig. 2. The following equations hold:

$$V=Ri_R=q_c/C, \tag{1}$$

$$E=R_0(i_R+i_c)=R_0 i, \tag{2}$$

which, for $t>0$ and constant i, are simply solved as

$$V=iR[1-e^{-t/RC}] \tag{3}$$

$$E=iR_0=\text{constant}. \tag{4}$$

If i is not constant, we still have the following relation by substituting Eq. (1) into Eq. (2).

$$E=R_0[(V/R)+(CdV/dt)]. \tag{5}$$

With an integrating factor $e^{t/RC}$, this integrates to

$$V=\int_0^t \frac{E(t')}{R_0C}\exp[-(t-t')/RC]dt'. \tag{6}$$

Since the experiment gives E as a function of t, we can determine V as a function of t. However, $E=iR_0$, so we can also determine i as a function of V.

Although it may be possible to write down a complete expression for the transient response, which includes the effect of initial velocity of the secondaries, space charge limitation of secondaries, and target leakage, it is not likely that significant experimental results could be obtained when all factors are equally important. Consequently, certain limiting cases will be considered which will give much more insight into the problem.

If the collector voltage V_c is sufficiently high, the leakage through the target will limit the ultimate potential of the target to a value less than V_c and the velocity distribution of the emitted secondaries will not play any observable role. On the other hand, if V_c is sufficiently low, the ultimate target potential may be practically entirely determined by the velocity distribution and the leakage can be neglected. There is also the intermediate case in which the ultimate target potential exceeds V_c but is in part determined by the leakage current and in part by the velocity distribution of secondaries. In all cases we shall assume that the primary current density is sufficiently low that there is no appreciable space charge limitation of the emitted secondaries. We shall also assume that the collector potential V_c is very much larger than the maximum value of E.

Case I. Ultimate Surface Potential is Limited Solely by Target Leakage

If we assume that the yield is not a function of the field across the target for an accelerating collecting field, i.e., no field enhanced emission, we have the following:

$$i=j_pA(\delta-1)$$
$$C=0.0885\kappa Al^{-1}$$
$$R=\rho l A^{-1}.$$

Substituting these into Eqs. (3) and (4) we obtain

$$V=j_p\rho l(\delta-1)[1-\exp(-t/0.0885\kappa\rho)] \quad (7)$$

$$E=j_pAR_0(\delta-1), \quad (8)$$

when

$$t>0 \quad \text{and} \quad V<V_c.$$

The observed voltage pulse is a constant unless there is field enhanced emission. If the latter exists, it will show directly on the observed voltage E and the field required to affect it can be determined from Eq. (7).[4]

Case II. Ultimate Surface Potential is Limited Solely by Collector Potential V_c

Again we assume that there is no field enhanced emission and also that the leakage resistance $R=\infty$. C is as defined above.

The integration now consists of two parts:
(1) $0<V<V_c$. Just as in Case I, we have

$$i=j_pA(\delta-1) \quad (9)$$
$$E=j_pAR_0(\delta-1).$$

To obtain V, we expand the exponential in Eq. (7) and let $\rho\to\infty$ from which we obtain

$$V=lj_p(\delta-1)t/0.0885\kappa. \quad (10)$$

This holds up to

$$t=\tau=0.0885\kappa V_c/lj_p(\delta-1). \quad (11)$$

[4] K. G. McKay, Advances in Electronics 1, 109 (1948).

(2) $V>V_c$ or $t>\tau$. Here we have

$$i=j_pA\left[\delta-1-\int_{V_c}^{V}P(V-V_c)dV\right],$$

where $P(V-V_c)=P(E_s)=$ the probability energy distribution function of the secondary electrons.

$$E=R_0j_pA\left[\delta-1-\int_{V_c}^{V}P(V-V_c)dV\right] \quad (12)$$

or

$$dE/dt=-R_0j_pA(dV/dt)P(V-V_c). \quad (13)$$

From Eq. (5) with $R\to\infty$

$$dV/dt=El/0.0885\kappa AR_0. \quad (14)$$

Substituting this in Eq. (13) and rearranging, we obtain

$$P(E_s)=-(0.0885\kappa/j_plE)(dE/dt), \quad (15)$$

where $E_s=V-V_c$. From the experiment we obtain E and dE/dt as a function of time. Equation (15) then gives $P(E_s)$ as a function of time. From Eq. (6)

$$E_s=V-V_c=\int_0^t(Eldt/0.0885\kappa AR_0)-V_c. \quad (16)$$

Thus knowing E from the experiment as a function of time, we can determine E_s as a function of time. Using t as a parameter, we can plot $P(E_s)$ as a function of E_s. This will serve to determine the energy distribution of secondary electrons up to the time when the effective yield falls to unity, which corresponds to the voltage V for which the term in brackets in Eq. (12) goes to zero. The velocity distribution of high energy secondaries will not be measured, but this may not be too important for a high yield surface.

Case III. Ultimate Surface Potential is Limited Both by Leakage and by Collector Potential

This is the intermediate case for which the target potential exceeds V_c but the ultimate effective yield is greater than unity. For $V<V_c$ Case I applies and the expressions for V and E are given by Eqs. (7) and (8). These hold up to $V=V_c$ or $t=\tau$, where τ is now given by the solution of Eq. (7) as

$$\tau=0.0885\kappa\rho\cdot\ln[1-V_c/j_p\rho l(\delta-1)]. \quad (17)$$

For $t>\tau$ we have E given by Eq. (12) in Case II. However, from Eq. (6)

$$V=\int_0^t\frac{E(t')\exp[-(t-t')/0.0885\kappa\rho]}{0.0885\kappa AR_0/l}dt', \quad (18)$$

which gives V as a function of t. From Eq. (5)

$$dV/dt=(l/0.0885\kappa A)[(E/R_0)-(VA/\rho l)]. \quad (19)$$

Substituting Eq. (19) into Eq. (13) and rearranging,

we obtain

$$P(E_s) = -(0.0885\kappa/j_p l) \times [E-(AR_0/\rho l)(E_s+V_c)]^{-1} dE/dt. \quad (20)$$

As before, $E_s = V - V_c$, and this can be determined as a function of time from the experimental data by means of Eq. (18). Equation (20) expresses $P(E_s)$ as a function of time and hence, if t is used as a parameter, $P(E_s)$ can be obtained explicitly as a function of E_s. The analysis is more complex than in Case II and the accuracy is not as good.

4. SUBSIDIARY MEASUREMENTS

The actual values of ρ and κ are not always known for the target material. However, we can determine τ, the time taken for the surface potential to equal V_c, since there is a change in the slope of E at this point. We can determine τ for two different values of V_c and use these in conjunction with Eq. (17) to solve for κ and ρ. If a low value of collector voltage is used, or if ρ is high, we may be able to neglect the leakage and obtain κ directly from Eq. (11).

5. SPACE CHARGE LIMITATIONS

Throughout this discussion we have assumed that the emitted current is not limited by space charge. This assumption will be true if the absolute value of the actual potential in the target-collector space, measured relative to the target, is everywhere less than the absolute value of the retarding potential. The effect of space charge will be most pronounced for low values of retarding field; i.e., $V \simeq V_c$. This will make the break in the observed voltage at $t=\tau$ less sharp and tend to over-emphasize the number of slow secondaries. However, a reasonable signal-to-noise ratio can be obtained with currents sufficiently small to render this effect almost negligible.

Another effect due to space charge has been pointed out by Williams and Kilburn.[5] If a pulsed beam of electrons strikes an insulating target which is in a retarding field such as to make the effective yield equal to unity, there will be, of course, no net current flow to the target surface during the pulse. However, the mere presence of the retarded secondaries in the space in front of the target induces a positive charge on the target backing place. This appears experimentally as an impulse excitation of the amplifier at the beginning of the pulse and another in the reverse direction at the end of the pulse. With varying intensity this will occur for any retarding field irrespective of the effective yield. For this reason, the energy distribution of secondaries emitted from an insulator is difficult to determine merely by measuring the initial yield from a target which is in a retarding field when the bombardment begins, since the initial yield has this impulse excitation superimposed on it.

[5] F. C. Williams and T. Kilburn, Proc. Inst. Elec. Engrs., Pt. III 96, 81 (1949).

Let us see how this effect may influence the transient response considered in Sec. 3. It can be readily shown that under any normal circumstances such an effect is completely negligible when the target is in an accelerating collecting field. However, as the target surface potential increases so as to exceed the collector potential, the density of electrons in the space between the target and the collector will increase and a corresponding charge will flow through the load resistor R_0. This will take place gradually rather than abruptly as in Williams' case, so the voltage amplitude of the disturbance will be correspondingly reduced. It appears probable that if the current densities are low enough to produce negligible space charge fields by the criterion of the preceding paragraph, they will also be low enough to prevent any appreciable modification of the transient wave form by this image charge. The only part of the wave form which is likely to be modified is the trailing edge where the bombarding beam is cut off. This may exhibit an impulse in the direction corresponding to electrons striking the target, but this portion is not of interest in the analysis.

6. EFFECT OF PULSING THE COLLECTOR

It is also possible to pulse the collector voltage and, in theory, obtain additional information. As an example, let us assume that the target has essentially reached its final equilibrium potential under bombardment for Case II where $\rho = \infty$. The effective yield is now unity. We abruptly decrease the collector potential V_c to zero. The effective yield becomes less than unity and the target potential will decrease until unity effective yield is again obtained. The previous analysis is valid except for sign changes, and thus the range over which $P(E_s)$ can be determined can be extended by an amount of the argument equal to the change in V_c. However, when V_c is changed abruptly to zero, the capacity between the collector and the target with its supports is sufficient to introduce a large transient into the input circuit. Under normal conditions this will overload the amplifier and consequently some form of bucking circuit would have to be used. This can be done, but it is not easy to accomplish it without losing part of the transient which is to be measured.

The problem of the induced transient does not arise if the collector voltage is changed outside of the measuring time interval provided that sufficient time elapses after such a change to permit the amplifier to recover before the onset of the next pulse to be measured. This elapsed time interval can be greatly decreased by the use of a crude form of bucking circuit. Now, instead of using collector pulsing to obtain more information about the energy distribution, we use it to restore the target potential to zero. Suppose that at the end of the measuring pulse, the target has reached its equilibrium value of $V_c + V(\infty)$, where the effective yield is unity. The collector voltage is dropped to $-V(\infty)$ and bombardment continues until equilibrium is again reached. The

target surface will now be at ground potential. Bombardment is stopped, the collector voltage is restored to a potential V_c, and the target is now ready for the onset of the next measuring pulse.* There are many variations of this sequence which are possible; and, where adequate bucking circuits are available, the functions of target potential restoration and extension of the energy distribution measurement can be combined.

7. GENERAL PROBLEMS RELATED TO ENERGY DISTRIBUTION MEASUREMENTS

In the foregoing analysis no mention has been made of the part played by the work functions of the target and of the collector nor of the effect of patch fields on the surface of the target. In the absence of bombardment, the equilibrium potential between target and collector will not be exactly that determined by the applied voltage on the collector but will be modified by the work functions of the collector and of the target. The work function depends on the position of the Fermi level relative to the energy bands, and in an insulator the Fermi level may fluctuate widely because of the existence of relatively small densities of impurities or trapped charges.[6] This introduces an uncertainty in the value of the surface potential which, in a polycrystalline or contaminated target would be aggravated by variations in surface potential over the target area. This suggests that during the interval between pulses, the neutralization of surface charges to produce a known surface potential should be accomplished by bombardment of the surface by low energy electrons rather than by relying on volume conductivity. It is possible that the former method should produce a more uniform surface potential. These considerations suggest that no matter what method of measurement is used, the accuracy of the determination of the velocity distribution of secondary electrons emitted from insulators may never equal that obtainable from metallic targets.

8. NUMERICAL EXAMPLE

It is instructive to consider the values of the various parameters which might reasonably be encountered in an actual experiment and to analyze a representative wave form. This not only establishes the order of magnitude involved but also evaluates the accuracy with which the mathematical analysis can be performed.

Let us assume that we have a video amplifier with a good noise figure and a band width of 20 mc/sec. The band width, together with the input capacity, sets an upper limit on the value of R_0. The voltage developed across R_0 by the primary bombarding current should be at least five times the noise level and this sets a lower limit on i_p. This value should be approached to minimize space charge effects. The target dimensions and the

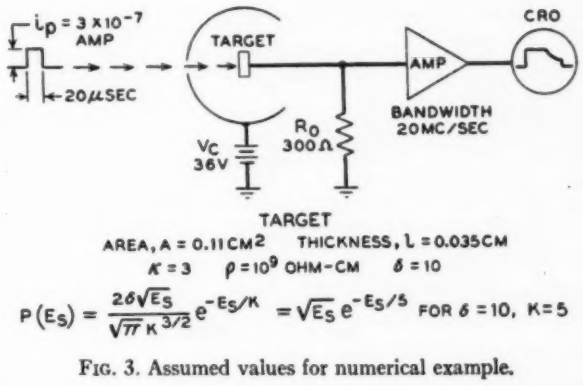


FIG. 3. Assumed values for numerical example.

value of the collector voltage should be such as to include all of the interesting part of the voltage transient on a time scale which is compatable with the resolution time of the amplifier. We shall concentrate on Case II in which leakage is negligible. Bearing these factors in mind we have picked as suitable values those shown in Fig. 3.

Rather than use an experimentally obtained transient response, we shall assume an energy distribution of the secondaries and from that derive the anticipated transient response. This calculated transient response will be used as "experimental" data and analyzed to produce an energy distribution which is to be compared with the originally assumed distribution. This approach has the advantage that it enables us to make a check on the accuracy of the mathematical analysis that can be performed on a graphically displayed transient. Let us assume that the energy distribution of the secondaries is Maxwellian. Such a distribution function does not describe properly the distribution of high energy secondaries, but that is irrelevant since we cannot determine their distribution by this method. The normalized distribution function which we assume is thus

$$P(E_s) = 2\delta E_s^{\frac{1}{2}} e^{-E_s/K}/\pi^{\frac{1}{2}}K^{\frac{3}{2}} = E_s^{\frac{1}{2}} e^{-E_s/5} \qquad (21)$$

for $\delta = 10$ and $K = 5$ volts.

Replacing E_s by $V - V_c$ we substitute Eq. (21) into Eq. (12), which can be integrated explicitly to yield E as a function of V. Substituting this result into Eq. (14) and integrating numerically we can obtain V as a function of t and from the previously obtained relation,

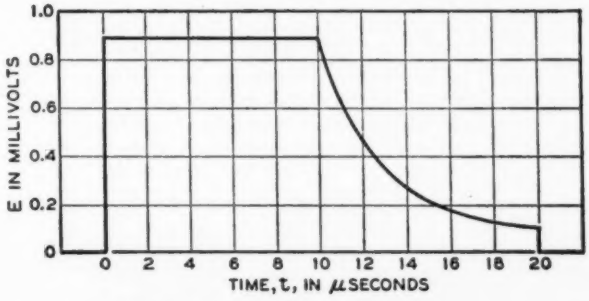


FIG. 4. Calculated transient response.

* A similar arrangement has been described by H. L. Heydt, Rev. Sci. Instr. 21, 639 (1950).

[6] G. H. Wannier, Phys. Rev. 76, 438 (1949).

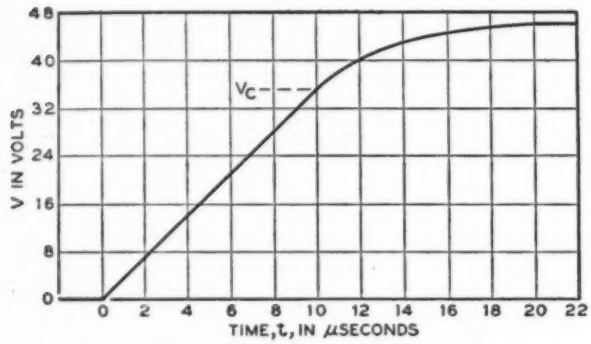


FIG. 5. Calculated variation of potential of target surface with time.

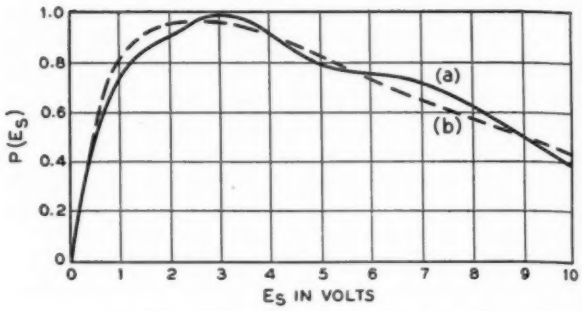


FIG. 6. (a) Calculated energy distribution of secondaries. (b) Originally assumed energy distribution.

determine E as a function of t, which is the desired transient. Figure 4 shows this calculated transient using the values assumed above. The break at the point where $V = V_c$ is quite pronounced. This transient curve was treated as "experimental" data. It was given to a computer who, without knowledge of the steps by which it had been obtained, was instructed to analyze it according to Eqs. (15) and (16) so as to yield an energy distribution. Figure 5 shows the calculated variation of target potential V with time and Fig. 6 is the calculated energy distribution with the originally assumed distribution shown as a dotted line. It will be seen that the accuracy leaves something to be desired. Probably the errors involved in deriving the transient are no greater than those that would normally be introduced by experimental factors. However, the peak of the distribution has been located within 0.5 volt and, when we ignore the fluctuations in the calculated curve, a still better reproduction of the assumed distribution is obtained.

8. CONCLUSIONS

It has been demonstrated that the transient response obtained from an insulator bombarded by a pulse of electrons can give not only a value for the yield but also values for the resistivity, the dielectric constant, and part of the energy distribution function of the secondaries. If a high yield target with negligible leakage is used, most of the energy distribution function can be obtained. The accuracy of the method is not high since both integration and differentiation of an observed wave form are involved. However, in many instances even a crude knowledge of the energy distribution would be valuable. Where leakage is negligible, all of the necessary experimental data are contained in one photographed transient. Thus, when the energy distribution may be changing as a function of time owing to changes in the target, this can readily be followed by this method. Another interesting aspect is that the lower the leakage, the more accurate are the results. This contrasts with most methods used previously where the insulator was made temporarily conducting by using thin films, by heating or by irradiation with infrared, etc.

I wish to express my appreciation for many fruitful discussions of this work with C. Herring, J. B. Johnson, and A. H. White.

Books Received

Instruments and Accessories for Radio-Isotope Applications

The Scientific Instrument Manufacturers Association of Great Britain has published a Brochure of Instruments and Accessories for Radio-Isotope Applications. This publication has been produced in collaboration with A.E.R.E. Harwell and is intended to summarize the equipment available from British sources for Isotope Techniques in medicine, research, and industry.

Copies of this publication are available free on application to the office of The Scientific Instrument Manufacturers Association of Great Britain Ltd., 17 Princes Gate, London S.W. 7, England.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 1 JANUARY, 1951

Microwave Measurements of the Dielectric Properties of Gases*

GEORGE BIRNBAUM, S. J. KRYDER, AND HAROLD LYONS
National Bureau of Standards, Washington, D. C.

(Received August 9, 1950)

An accurate and sensitive method for measuring the complex dielectric constant ($\epsilon'-i\epsilon''$) of gases in the microwave region is described and critically investigated for sources of error. A resonant-cavity method is used in which the cavity response curve is displayed on a cathode-ray tube. The variation of resonance frequency and Q of the cavity when filled with gas are determined from measurements with two frequency markers. The equipment is operated at 9000 Mc, and cavities with Q's of 1×10^4 are used. In the determination of $\epsilon'-1$ (dimensionless) of very low loss gases, an accuracy of 0.4 percent and a sensitivity of 4×10^{-7} can be obtained. In the determination of the loss factor ϵ'', an accuracy of 2 percent and a sensitivity of 5×10^{-7} can be obtained. This method is useful for measuring dispersion and absorption in solids and liquids as well as gases.

Experimental results are given and briefly discussed for $\epsilon'-1$ of O_2, N_2, CO_2, He, and air, and for ϵ'' of NH_3 as a function of pressure up to 20 cm Hg.

IN Part I of this paper a resonant-cavity method for measuring the dielectric properties of gases is described and critically examined for sources of error. In Part II results obtained by this method at 9000 Mc are given and briefly discussed. These results include absorption in NH_3 and the dielectric constants of five nonpolar gases.

PART I. MEASUREMENT TECHNIQUE

A. Introduction

A resonant-cavity method was used by Saxton[1] for measuring the dielectric properties of water vapor at 9 cm, 3.2 cm, and 1.6 cm, early in the development of microwave techniques. The variations of the resonance frequency and the Q-factor of the cavity when filled with vapor were determined with a detector and galvanometer. This method gave an accuracy of approximately 1×10^{-4} for $\epsilon'-1$, and a sensitivity of approximately 1×10^{-6} for the loss factor,[2] ϵ''. Greater sensitivity in the measurement of $\epsilon'-1$, namely, 1×10^{-5}, was reported by Jen[3] who devised a more elaborate cavity method in which a swept frequency was applied to the cavity, and the changes in resonance frequency and Q were determined from measurements on a cathode-ray tube (CRT). For work with relatively high susceptibility and high loss gases, Hershberger[4] employed a wave-guide system, less sensitive than the cavity methods. In addition to these methods, which measure both the susceptibility and the loss factor,[5] many specialized techniques have been developed to investigate the microwave spectra of gases,[6] and a few methods have been developed which measure just the susceptibility of gases.[7,8]

The present work was begun in 1946 and was undertaken to develop a more sensitive and accurate method for measuring the complex dielectric constant of gases in the microwave region at pressures where the absorption line is quite broad. In addition to the data reported in Part II this method has been used to study dispersion[9] in NH_3 and the dielectric properties of water vapor.[10]

A cavity method is used in which the cavity resonance curve is displayed on a CRT by a sweep-frequency technique. A feature in the method for measuring $\epsilon'-1$ is the use of two cavities: one contains the gas and the other acts as a frequency reference. This arrangement permits the frequency change of the test cavity due to the admission of gas to be measured directly by aligning a pair of frequency markers, in the form of intensified spots, at the peaks of the resonance pulses. By aligning these markers at the half-power points of the resonance curve, the band width and hence Q of the cavity can be determined. Another feature of this method is that it can be used to measure the complex dielectric constants of liquids and solids.[11]

B. Theory of Measurement

It is assumed that a matched generator and load are connected to a transmission-type cavity which operates at a high Q mode well separated from other modes.[12]

* Preliminary accounts of this work have been given by G. Birnbaum at the Washington, D. C., meeting of URSI-IRE, May, 1947; and by Lyons, Birnbaum, and Kryder, Phys. Rev. 74, 1210 (1948).

[1] J. A. Saxton, "The dielectric properties of water vapor at very high radio frequencies," *Meteorological Factors in Radio-wave Propagation* (The Physical Society, London), p. 215.

[2] The complex dielectric constant is given by $\epsilon=\epsilon_0(\epsilon'-i\epsilon'')$, where ϵ_0 is the dielectric constant for free space, its value depending upon the units used, ϵ' is the familiar value of dielectric constant relative to free space, and ϵ'' is called the loss factor.

[3] C. K. Jen, J. Appl. Phys. 19, 649 (1948).

[4] W. D. Hershberger, J. Appl. Phys. 17, 495 (1946).

[5] A review of the measurement of dielectric properties, especially of solids and liquids, and a fairly extensive bibliography of work in this field is given in C. G. Montgomery, *Techniques of Microwave Measurements*, MIT Radiation Laboratory Series, Vol. 11 (McGraw-Hill Book Company, Inc., New York, 1947).

[6] These methods have been reviewed by W. Gordy, Revs. Modern Phys. 20, 668 (1948).

[7] G. Birnbaum, Rev. Sci. Instr. 21, 169 (1950).

[8] C. M. Crain, Phys. Rev. 74, 691 (1948).

[9] G. Birnbaum, Phys. Rev. 77, 144 (1950).

[10] G. Birnbaum and S. K. Chatterjee (to be published).

[11] G. Birnbaum and J. Franeau, J. Appl. Phys. 20, 817 (1949).

[12] It is also necessary that (1) the direct mutual impedance of the cavity couplings is zero; and (2) the coupling losses are zero.

The power P which flows out through the output line is given by [13,14]

$$P/P_1=[1+4Q_1^2(f-f_1)^2/f_1^2]^{-1}, \quad (1)$$

where f_1 is the frequency at which maximum power P_1 is transmitted by the cavity, and $2(f-f_1)/f_1$ is the approximation for $[(f/f_1)-(f_1/f)]$. With negligible error, f_1 is also the frequency of the free oscillations of the lossless cavity. When $P/P_1=\frac{1}{2}$, the familiar relation for loaded-Q,[15]

$$Q_1=f_1/\Delta f_1, \quad (2)$$

is obtained, where Δf_1 is the bandwidth of the cavity at the half-power points.

When the cavity is filled with dielectric material whose complex dielectric constant is $\epsilon'-i\epsilon''$, the change in resonance frequency from f_1 to f_2 and the change in Q from Q_1 to Q_2 are given by

$$(f_1-f_2)/f_2=\epsilon'^{-\frac{1}{2}}-1, \quad (3)$$

and

$$\epsilon''=(1/Q_2)-(1/Q_1). \quad (4)$$

The variations of $\epsilon'-1$ and ϵ'' with frequency in a frequency interval approximately equal to the band width of the cavity are taken to be negligible. Another form of Eq. (4) will be found useful. Since the power transmitted by the cavity at resonance varies as Q^2 giving $Q_1/Q_2=(P_1/P_2)^{\frac{1}{2}}$, we have

$$\epsilon''=[1/Q_1][(P_1/P_2)^{\frac{1}{2}}-1]. \quad (5)$$

An idea of the values of the frequency differences which appear in these equations may be obtained from the following example. If $Q_1\simeq 1\times 10^4$, $f_1\simeq 1\times 10^4$ Mc, $(\epsilon'-1)\simeq 1\times 10^{-3}$, and $\epsilon''\simeq 1\times 10^{-4}$, then $(f_1-f_2)\simeq 5$ Mc, $\Delta f_1\simeq 1$ Mc, and $\Delta f_2\simeq 2$ Mc. In the equipment to be described, these frequency differences are measured directly.

Equation (3) is correct insofar as the losses in the dielectric can be neglected. The correction to f_2 taking into account these losses is of the order $f_2(\epsilon''/\epsilon')^2$, and is usually negligible for gases.

Q_1 depends, in part, on the unloaded Q of the cavity, Q_u, which varies as $1/f^{\frac{1}{2}}$. Actually, Q_1 should refer to the cavity-Q when the cavity contains a hypothetical sample with dielectric constant ϵ' but zero loss. Since with gases the frequency change is very small, the error in taking Q_1 equal to the Q of the empty cavity is negligible.

It should be noted that the Q and the frequency of maximum transmission depend upon the admittance of the generator and the load. Let the normalized admittance of the generator be g_g+ib_g and that of the load be g_l+ib_l (at some reference planes in the input and output transmission lines). These admittances, which are taken to differ from the characteristic admittance of the wave guide by a small amount, perturb f_1 to f_1' and Q_1 to Q_1':[14]

$$f_1'=f_1+(f_1/2Q_e)(b_l+b_g), \quad (6)$$

$$1/Q_1'=(1/Q_u)+(1/Q_e)(g_l+g_g). \quad (7)$$

In these equations Q_e is the Q due to the loss coupled into the cavity when the transmission line is matched. For convenience, it is assumed that the cavity couplings are of equal strength. What are usually known in microwave circuits are the voltage standing-wave ratios (VSWR's) of the generator and the load admittance. Depending on the phases of these VSWR's at the reference planes, f_1' and Q_1' can have a maximum or a minimum value. For a single line coupled to the cavity, the extreme values in f_1' correspond to $b=\pm(r^2-1)/2r$ and the extreme values in Q_1' correspond to $g=r^{\pm 1}$, where r is the VSWR. For the two-line cavity, the extreme values are

$$f_1'=f_1\pm\frac{f_1}{4Q_e}\left(\frac{r_g^2-1}{r_g}+\frac{r_l^2-1}{r_l}\right), \quad (8)$$

and

$$1/Q_1'=(1/Q_u)+(1/Q_e)(r_g^{\pm 1}+r_l^{\pm 1}). \quad (9)$$

However, the measured values of ϵ'' and $\epsilon'-1$ will be independent of the generator and the load admittance provided these admittances remain constant when the frequency is changed from f_1 to f_2.

C. Description of Equipment

The arrangement of the apparatus[16] is shown schematically in Fig. 1 and in the photograph, Fig. 2. A sawtooth generator simultaneously supplies a sweep voltage to an oscilloscope and a modulation voltage to the reflector of a klystron oscillator. The FM output of the klystron is applied to two cavities whose outputs are rectified, amplified, and displayed on a CRT. By means of an electronic switch, the two resonance curves can be presented on separate traces. A simple electronic circuit is used to control the polarity of one of the resonance pulses.

The FM oscillator also serves as the local oscillator for a heterodyne receiver whose intermediate frequency is adjustable. Since the FM oscillator varies above and below the microwave continuous wave (c-w) oscillator, the signals produced by the i-f amplifier consist of two pulses whose frequency separation is twice the intermediate frequency. After rectification, these pulses intensity modulate the CRT, thus producing two bright

[13] Montgomery, Dicke, and Purcell, *Principles of Microwave Circuits*, MIT Radiation Laboratory Series, Vol. 8 (McGraw-Hill Book Company, Inc., New York, 1948), Chapter 7.

[14] J. C. Slater, *Microwave Electronics* (D. Van Nostrand Company, Inc., New York, 1950), Chapters 4 and 5.

[15] Unless otherwise indicated, Q refers to the total or loaded value.

[16] This equipment, which combines the features of a cavity Q-meter and cavity comparator, was developed by modifying a cavity Q-meter. Circuit details of these equipments are given in reference 5. For a description of somewhat similar equipment see R. A. Unterberger and W. V. Smith, Rev. Sci. Instr. 19, 580 (1948).

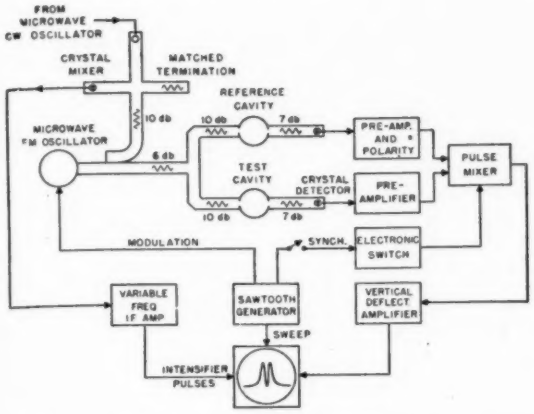


FIG. 1. Schematic diagram of electronic and microwave components used in measuring the dielectric properties of gases.

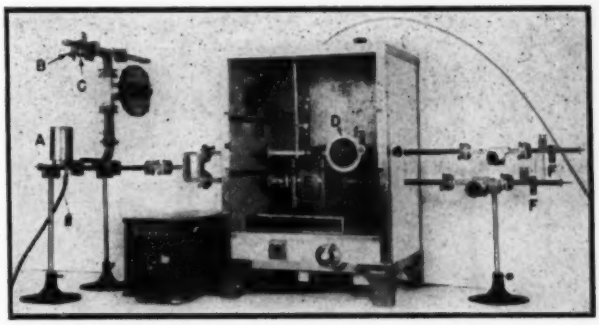


FIG. 2. General arrangement of equipment. A—FM oscillator; B—to c-w oscillator; C—to i-f amplifier; D—reference cavity; E—evacuation chamber, test cavity inside; F—to preamplifier.

spots on the CRT trace. By tuning the i-f amplifier, the frequency interval between these markers can be varied from 0.6 to 10 Mc. The band width of the i-f tuned circuit is high enough to give well-defined spots on the CRT trace, and low enough so that the pulses are not appreciably distorted or delayed. Hence, if it were desired to increase the selectivity of the i-f tuned circuit, it would be necessary to lower the present sweep rate of 60 cps. Even if there is some delay of the frequency markers, relative to the resonance pulses, no error results in the measurement of frequency difference with these markers unless the sweep voltage is not linear with time.

This is a versatile apparatus which can be used to measure Q, resonance frequency, and very small changes in resonance frequency of a cavity, and to compare Q's and transmission losses of two cavities. The equipment operates at 9000 Mc; by making suitable changes in the microwave components the operating range has been extended to the 3000-Mc and the 24,000-Mc region.

It is worthwhile noting the following details pertaining to the electronic circuits.[16]

1. The shape of the resonance pulse depended somewhat on the balance of the push-pull stage of the vertical deflection amplifier.
2. A small, spurious resonance was noted in the i-f amplifier at 3 Mc. This was easily damped by a resistor.
3. Appreciable reduction in the amplitude of hum on the oscilloscope trace was obtained by inserting an LC-filter in the plate-supply lead to the first stage of the vertical-deflection amplifier. It was then possible to reduce this hum to 0.1 in. at nearly full amplifier gain by destroying the path for ground currents via the wave guide. This was accomplished by insulating a wave-guide junction in each cavity circuit by placing a very thin sheet of mica across the wave guide.

The Microwave Circuits

Type 1N23A crystals in tunable mounts were used for the mixer and the detectors. A satisfactory match of the detector to the wave guide was obtained by tuning the crystal for maximum power, as indicated by maximum signal on the CRT. To make the latter adjustment with some accuracy, a small amount of attenuation was introduced between the cavity and the detector[17] in order to reduce the mismatch of the wave guide by the cavity. After tuning the detector for maximum power, the attenuation was increased to 7 db.

Attenuators (VSWR = 1.06) were placed between the cavity and the rest of the microwave circuits to provide the cavity with loads which were well-matched and insensitive to changes in frequency. The klystron was more poorly matched to the wave guide than the crystal detector; consequently, most of the attenuation was placed between the oscillator and the cavity. Frequency pulling of the oscillator by the cavity was noticed if the attenuation between these components was reduced below 12 db.

Frequency instability in the FM oscillator (a type 2K25 klystron) causes the frequency markers and the resonance pulses to shift simultaneously on the oscilloscope trace, while instability in the c-w oscillator causes relative motion between them. The latter instability is the more serious, since it decreases the precision in aligning the frequency markers. Two arrangements for stabilizing the frequency of the c-w oscillator have been used. In the simpler method, the factors which contributed to the frequency instability of a type 2K25 klystron were dealt with separately: the electrode voltages were supplied by batteries, and the tube was shielded from changes in room temperature and from mechanical vibrations. Such an arrangement gave a short-time frequency stability better than 1 part in 10^6, which proved to be adequate for most purposes. With the other frequency stabilization scheme, a Pound oscillator,[18] a short-time stability of about 1 part in 10^7 was realized. Fine tuning of the Pound oscillator was obtained by inserting the spindle of a micrometer a small distance into the stabilizing cavity of this oscillator. The spindle was silver-plated to prevent possible reduction in cavity-Q.

The Cavity Resonators

The cavities used in this work were standard, transmission-type wave meters, which operate in the TE_{011}

[17] Approximately 3 db was adequate to isolate the crystal from the cavity; more than this tended to make the tuning insensitive.
[18] R. V. Pound, Rev. Sci. Instr. 17, 490 (1946).

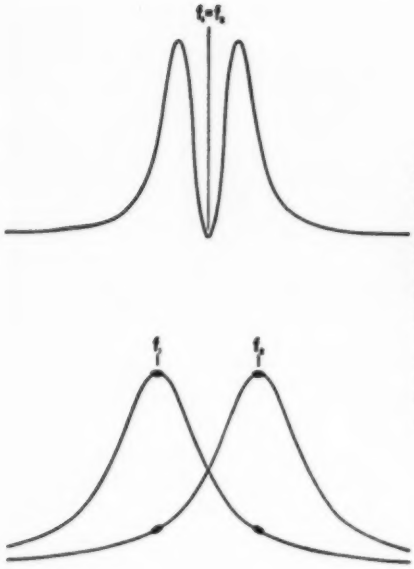


FIG. 3. (Upper) CRT pattern for frequency comparison; $Q_1=1.1Q_2$, $Q_1=1.0\times10^4$; vertical scale expanded. (Lower) CRT pattern for measurement of frequency difference; $(f_1-f_2)\backsimeq1.0$ Mc. For purposes of illustration, the size of the frequency markers is enlarged.

mode with a Q of about 1×10^4. At the frequency of operation, no evidence of spurious modes was noted in these cavities.[19] Mechanical distortion of the test cavity with evacuation was prevented by enclosing the test cavity in a metal chamber, equipped with gas and wave guide outlets.[20] The reference cavity was provided with a vernier tuning control, similar to the one used in the stabilizing cavity of the Pound oscillator.

The frequency difference between the reference and the evacuated test cavity must remain constant during the time required for the frequency change measurements. For cavities which are constructed from the same metal, this requires that the two cavities remain at the same temperature or temperature difference. In addition, the dielectric composition of the reference cavity should remain fixed. The first condition was realized by placing both cavities in a drying oven, which not only insulated them from fluctuations in room temperature but also provided a means for making measurements above room temperature. If the reference cavity is unsealed, its dielectric composition will depend upon that of the atmosphere; hence, its resonance frequency will vary significantly with small changes in the temperature, the pressure, and especially the water vapor content of the atmosphere. Since these changes were usually slow compared with the time of measurement, it was possible to use a partially sealed reference cavity.

D. The Measurement of Dielectric Constant

The determination of dielectric constant involves the measurement of the resonance frequency of the test cavity and its change in frequency due to the admission of gas. In a typical determination, these quantities were obtained in the following manner. After evacuating the test cavity, the resonance frequencies of the test and the reference cavity were equalized precisely by a method described by Sproull and Linder.[21] This method consists in superimposing on the CRT trace two resonance curves of opposite polarities. A typical pattern obtained for two cavities with nearly equal Q's and whose resonance curves were of nearly equal amplitudes is shown in Fig. 3. At this point, the frequency of the test cavity was obtained from the calibrated tuning control of the reference cavity. The calibration was known with an error of 0.01 percent, which is negligible in comparison with the error in measuring the frequency change. When gas was admitted into the test cavity, the superimposed resonance pulses separated. To measure this separation, both resonance pulses were made to have the same polarity, and then the frequency markers were aligned at the peaks of the resonance curves, as shown in Fig. 3. The resonance curves were placed on separate traces, by the electronic switch, to avoid the slight shift in the peak of each resonance curve due to the tail of the other. In order to observe small motions of the frequency markers, the CRT scale was adjusted for a high dispersion in in./Mc. Finally, the gas was pumped out of the test cavity and the initial or zero adjustment was checked.

This procedure was modified for measuring frequency changes less than 0.6 Mc, lower limit of the tuning range of the i-f amplifier. In this case, the cavities were initially displaced in frequency by at least 0.6 Mc, and this displacement was accurately measured by the markers.

Sources of Error

1. *Measurement of frequency change.*—Assuming that the patterns of the CRT have been adjusted to best advantage, the precision in measuring frequency change depends upon the Q's of the cavities. The following results were obtained at 9000 Mc with cavities whose Q's are equal to 1×10^4. In the manner described previously, the resonance frequencies of two cavities were equalized with an error of 0.002 Mc. The precision in measuring frequency difference with the markers depends also on the frequency stability of the c-w oscillator. With the battery-stabilized klystron the precision was 0.5 percent, and with the Pound oscillator the precision was improved to 0.002 Mc or 0.2 percent, whichever was greater. Except where otherwise indicated, all data given in this paper were obtained when the Pound oscillator was in use.

The tuning scale of the i-f amplifier was calibrated by a signal generator with an error of 0.3 percent. This error was due mainly to the uncertainty in the calibration of the signal generator.

2. *Effect of the cavity loads.*—Some VSWR measurements were made on the detector system (the attenuator and detector) primarily to obtain an estimate of the

[19] A discussion of mode crossing and its effect on resonance frequency and Q of a cavity is given by Bleaney, Loubser, and Penrose, Proc. Phys. Soc. (London) 59, 185(1947).

[20] Reference 7, Fig. 6.

[21] R. L. Sproull and E. G. Linder, Proc. Inst. Radio Engrs. 34, 305 (1946).

error introduced in a frequency change measurement by changes in the susceptances of the cavity loads. With the detector system matched to the wave guide as described previously, it was found that the VSWR changed from 1.08 to 1.09 when the frequency was changed by 10 Mc. Assuming that these VSWR's exist on both sides of the cavity, we find from Eq. (8) that if $Q_e = 4\times10^4$, the maximum error in measuring a frequency change of 10 Mc could be only 0.02 percent, which is negligible compared with other errors. Taking into account the VSWR's of the mica windows used to seal the test cavity does not significantly change this result.

3. *Shape of the resonance curve.*—Asymmetry in the resonance pulses will cause an error, in general, in the frequency comparison of two cavities and in the measurement of frequency difference. The factors which can contribute to this asymmetry include: (1) distortion of the cavity response due to excitation by a frequency varying with time; (2) distortion by the amplifier; (3) amplitude modulation of the FM oscillator; and (4) hum on the CRT trace. The first factor is thought to be negligible;[22] the second has been practically eliminated; and the third can be avoided as follows. As is well known, the power output from a klystron varies with reflector voltage, except for a small region at the peak of the klystron's voltage mode. Hence, the tuning of the klystron should be adjusted so that the resonance frequency of the cavity coincides with the constant power region.

The effect of hum on the accuracy of frequency comparison and frequency difference measurements depends not only upon its amplitude, but also upon its shape. As an example of the magnitude of this effect, suppose that a voltage which rises linearly on the CRT is superimposed on the resonance pulse. Then a simple calculation shows that if this voltage rises 0.1 in. per Mc, the amplitude of the resonance pulse is 2 in., and the cavity band width is 1 Mc, then the peak of the resonance pulse will be shifted by 0.006 Mc. If this linear voltage were added to the CRT pattern obtained when two resonance pulses of opposite polarity were superimposed (see Fig. 3), the heights of the shoulders would be unequal; hence, an erroneous frequency difference between the cavities would be indicated. In the present equipment, the hum on the oscilloscope trace was reduced to 0.1 in., and it is thought that the error introduced in a frequency change measurement by this hum is no greater than 0.004 Mc.

4. *Temperature stability of the cavities.*—For a homogeneous cavity undergoing thermal expansion, the fractional change in frequency due to a change in temperature is equal to $-\alpha$ per °C, where α is the coefficient of linear expansion for the material.[5] In the present equipment it was noticed that over a 15-minute interval, the temperature difference between the two cavities, at room temperature, might fluctuate by 0.03°C. Since the cavities are constructed from steel for which $\alpha \simeq 1\times10^{-5}$ per °C, the frequency difference between the cavities might vary by 0.003 Mc. Of course, it is possible to correct for a steady change in temperature difference.

5. *Summary.*—The smallest measurable frequency change is determined by the precision in aligning the frequency markers as the ultimate limit. With the present equipment this value is 0.002 Mc, and from Eq. (3) the sensitivity in measuring $\epsilon' - 1$ is 4×10^{-7}. The total error in measuring frequency change is the sum of the aforementioned precision plus the errors discussed in the preceding sections. It is estimated from a survey of these errors and from experience with the apparatus that this total error may be taken as 0.008 Mc or 0.7 percent, whichever is greater. If the gas temperature and pressure are measured each with an error of 0.1 percent, the total error in measuring $\epsilon' - 1$ is increased to 0.9 percent. The corresponding rms error is 0.4 percent. These results, strictly speaking, apply for zero-loss gases.

Dielectric loss decreases the Q and the power transmitted by the cavity. Both effects reduce the precision in locating the peak of the resonance curve: the first by broadening the curve and the second by increasing the noise to signal ratio. It is in the latter case that lack of power in the FM oscillator is a serious handicap. However, the obvious expedient of reducing the attenuation in the wave-guide circuit can be used until frequency-pulling effects become serious. It is estimated that the rms error in measuring $\epsilon' - 1$ of a gas with a loss factor of 2×10^{-4} is approximately 1 percent.

When working with lossy gases, a cavity with an initial Q lower than 1×10^4, for example, can be used advantageously. As can be seen by inspecting Eq. (5), the decrease in power transmitted by a cavity containing a lossy dielectric is smaller for a smaller initial Q.

E. Loss Factor Measurements

Two methods have been used to determine the loss factor: one involves measuring the band width of the cavity at the half-power points as shown in Fig. 4, and the other involves measuring the peak amplitude of the resonance curve. Very small changes in this amplitude were determined by comparing the amplitudes of the two resonance curves. The width of the resonance pulse was adjusted so that it was clear that it decayed to zero at the ends of the sweep. This precaution was necessary because the baseline was a reference in determining the amplitude. The initial amplitude was set at 2 in. For accurate band-width measurements, it was essential that the CRT trace be sharp and practically free from noise and hum.

When a calibrated attenuator is available, the half-power points of the resonance curve can be found conveniently by inserting 3-db attenuation in the wave-

[22] G. Hok, J. Appl. Phys. **19**, 242 (1948). He defines a parameter α which may be written as $(\pi\Delta f)/2(\pi F\tau)^{\frac{1}{2}}$, where Δf = band width of cavity, F = maximum frequency deviation, and τ = repetition rate. The distortion in the shape of the resonance pulse is quite small for $\alpha = 2$ (Figs. 2 and 8). In the present case $\alpha \simeq 50$.

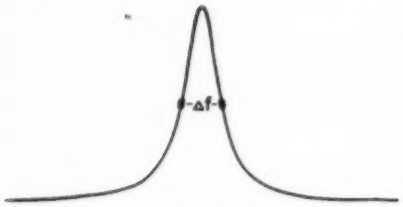


FIG. 4. CRT pattern for measuring cavity band width at the half-power points. For purposes of illustration, the size of the frequency markers is enlarged.

guide circuit. However, if this is to be done with accuracy, it is necessary that the attenuator be well matched on both sides.

Sources of Error

1. *Measurement of band width and amplitude.*—The precision in measuring the band width of the cavity in the neighborhood of the half-power points was found to be better than 1 percent for a range of Q's from 3×10^3 to 1×10^4. However, the error of the measurement was essentially that of establishing the half-power level. This error, due to the combined effect of the finite trace width, noise, and hum, was approximately 2 percent. Since the error in determining resonance frequency was negligible, the rms error for a Q measurement was approximately 2 percent. From similar considerations, it was estimated that the rms error in measuring the peak amplitude of the resonance curve was 0.03 in.

2. *Shape of the resonance curve.*—The accuracy of the loss factor measurements depends upon the assumption that the form of the resonance curve, as measured with the frequency markers, can be represented by Eq. (1). Since the power transmitted by the cavity is given by Eq. (1), it is necessary that: (1) the vertical amplifier pass undistorted the signals applied to it; (2) there be no amplitude modulation of the FM oscillator; and (3) the crystal detector operate as a square-law device. It is also convenient to require that the klystron frequency be directly proportional to reflector voltage, so that an oscilloscope scale linear in frequency is obtained. This condition was realized by operating in the constant power region of the klystron.

The first and second conditions have been discussed in other parts of this paper. However, the question of amplitude modulation of the FM oscillator needs further comment. Distortion in the shape of the resonance pulse due to this effect will be noticed when the bandwidth of the cavity becomes comparable to that of the mode curve of the klystron. However, it is easy to correct for this distortion if the relative power from the FM oscillator is known as a function of frequency. A Q less than approximately 500 cannot be measured with a sweep-frequency technique, since the range of frequency modulation obtained with the type 2K25 klystron is too small.

It was possible to show that the crystal detector was responsible for the residual distortion in the shape of the resonance pulse; hence, this fact was used to check the operation of the crystal. In practice, it was possible to select crystals such that the correct shape of the resonance curve and, consequently, a constant value of Q was obtained for several values of rf power.

If the shape of the resonance curve is correct, the value of Q calculated from band-width measurements made at various amplitudes of the resonance curve should be constant. The extent to which this was approached is illustrated in Table I. An alternative procedure was to plot P_1/P *vs* $(f-f_1)^2$ (see Eq. (1)). A straight line should be obtained if the resonance curve has the correct shape.

3. *Effect of the cavity loads.*—The error in measuring ϵ'', unlike the corresponding error in $\epsilon'-1$, could be serious when the cavity loads vary even slightly with frequency. Thus, by using the data given in Sec. D2 in Eqs. (4) and (9), it is found that the error in ϵ'' could be 5×10^{-7}, which is comparable with other errors of measurement.

4. *Summary.*—From the error in measuring Q (Sec. E1), it is calculated by the usual method of propagating errors that the rms error in measuring ϵ'' by the Q-method is 3×10^{-6} or 2 percent, whichever is greater. For small values of ϵ'', the error of the amplitude method is somewhat smaller. The sensitivity or smallest loss factor which can be detected by this method is approximately 5×10^{-7}. These values for the errors and sensitivities of the methods depend upon the initial value of Q.

The dielectric loss decreases the power transmitted by the cavity. In order to maintain the same fractional accuracy in locating the half-power points, the amplitude of the resonance curve must be increased to its original value. This is done, of course, by increasing the amplifier gain, although the increase in noise may reduce the accuracy of the band-width determination. In the present arrangement this decrease in accuracy is not significant until the loss factor is approximately 2×10^{-4}. To measure ϵ'' greater than this, the attenuation in the wave-guide circuit may be reduced until "Q-pulling" (Sec. E3) becomes serious, or a cavity with a lower initial Q can be used.

F. Discussion

The equipment and techniques described in this paper have been in use for several years and have been found to be reliable and convenient. The same techniques and electronic apparatus have been used to determine the dielectric properties of small samples of solids and liquids as well as gases.

TABLE I. The cavity-Q calculated from band-width measurements at various amplitudes.

Relative amplitude	Q
0.1	4840
0.2	4930
0.5	4960
0.8	4980
0.9	4920

A lossy gas can reduce the power transmitted by the cavity to the point where noise on the resonance pulse decreases the accuracy of the method. Techniques for dealing with this problem have been discussed. Another approach is to limit the gas to a small portion of the cavity volume. This can be accomplished by inserting in the cavity a tube made from quartz or from a low loss glass.

PART II. DIELECTRIC PROPERTIES OF O_2, N_2, CO_2 He, AIR, AND NH_3 AT 9000 MEGACYCLES

A. Introduction

The dielectric properties of NH_3 and some nonpolar gases were investigated at 9000 Mc to obtain data of interest as well as to demonstrate the accuracy of the methods of Part I. At this frequency, which is still far removed from the region of absorption in the nonpolar gases, the dielectric constant should not differ from the low frequency value. Hence, it is interesting to compare the present values with those obtained at low frequencies, especially since there appears to be discrepancies among the low frequency values of various observers.

Absorption in NH_3 was measured at low pressures on the low frequency side of the inversion spectrum at 24×10^3 Mc. The variation of absorption with pressure is compared with theory. Data on the dielectric constant of NH_3 at 3×10^3 Mc and 9×10^3 Mc are presented and discussed elsewhere.[9]

B. Dielectric Constants of O_2, N_2, CO_2, He, and Air

The gases used in this work were obtained from commercially available cylinders. These gases were analyzed for impurities by a mass spectrometer at the National Bureau of Standards.[23] It was found that, except for He, the impurities were so low in concentration as to have negligible effect on the measured value of $\epsilon'-1$. The 0.1 mole percent of N_2 in He makes the measured value of $\epsilon'-1$ for He too high by 0.5×10^{-6}; the value given for He in Table II was corrected by this amount. All the gases were dried in the process of filling the test cavity by being passed slowly through a tower containing anhydrous $CaSO_4$ followed by P_2O_5. Most of the measurements were made at room temperature and at a pressure of approximately 760 mm Hg. In computing $\epsilon'-1$ at 0°C, 760 mm Hg, the measured pressure was corrected by van der Waals' equation to the pressure that would be measured were the gas ideal.

TABLE II. Summary of experimental results: the dielectric constants of O_2, N_2, CO_2, He, and air, at 0°C and 760 mm Hg.

Gas	Present results[a] 9280 Mc $(\epsilon'-1)10^4$	Results of other observers $(\epsilon'-1)10^4$
O_2	5.300±0.019	5.233±0.010[b] 5.30[c] 5.325±0.013[d]
N_2	5.869±0.029	5.796±0.010[b]
CO_2	9.855±0.030	9.875±0.020[b]
He	0.705±0.011	0.684±0.005[b] 0.692±0.002[d]
Air	5.754±0.014	5.669±0.010[b] 5.72[c] 5.7520[e]

[a] The uncertainty is the rms value of approximately 15 determinations for each gas. The total error is larger.
[b] L. G. Hector and D. L. Woernley, Phys. Rev. **69**, 101 (1946). Measurements at 1 Mc.
[c] Reference 8. Measurements at 9370 Mc.
[d] J. G. Jelatis, J. Appl. Phys. **19**, 419 (1948). Measurements at 1 Mc.
[e] Extrapolated from the optical measurements of H. Barrell and J. E. Sears, Trans. Roy. Soc. (London) **238A** (1939–40).

[23] Mass Spectrometry Section, Division of Atomic and Radiation Physics, National Bureau of Standards.

TABLE III. The dielectric constant of dry air (0°C, 760 mm Hg): a comparison between the measured value and the value computed from its components.

Gas	Percent (partial pressure)	$(\epsilon'-1)10^4$	Contribution to air $(\epsilon'-1)10^4$
N_2	78.06	5.869	4.581
O_2	21.00	5.300	1.112
A	0.94	5.542	0.052
	Air: calculated from components given above	5.745	
	Air: measured	5.754	

The experimental results are given in Table II, which for purposes of comparison also gives the results of other recent investigations. In comparing the values given in Table II, it is interesting to note that Jelatis has shown excellent agreement between the values he obtained for the dielectric constant of some of the rare gases and the values extrapolated from optical measurements. The precision of these optical measurements is one or two orders of magnitude better than what is obtained with present day radio or microwave techniques. The present values are in close agreement with those of Jelatis. The values of Hector and Woernley are lower from 1 to 5 percent than those of Jelatis. Hector and Woernley, whose work precedes that of Jelatis, have given a survey of the literature which shows that considerable disagreement exists in the values obtained by various observers for the dielectric constant of many gases.

The uncertainty attached to the present results is the rms deviation from the average of approximately 15 determinations for each gas. The total rms error is estimated to be 0.7 percent for all gases except He. This is somewhat larger than the value given in Sec. I.D5 because the less stable c-w oscillator was in use when these measurements were made. The percentage error for He, namely, 2 percent, is larger because of its much lower value of susceptibility.

Thus far it has been tacitly assumed that the magnetic susceptibility of the gas is zero. When this does not hold, the fractional change in the resonance frequency of the cavity due to the admission of gas is given by

$$(f_1-f_2)/f_2=\tfrac{1}{2}(\epsilon'-1)+\tfrac{1}{2}(\mu'-1).$$

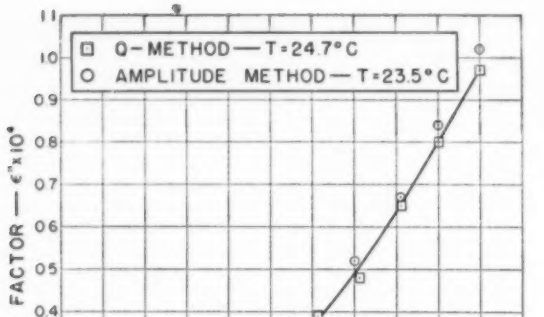


FIG. 5. Variation of loss factor ϵ'' with pressure in NH_3 at 9280 Mc. The curve is drawn through the ⊡-points.

For O_2, which is strongly paramagnetic, $\mu'-1$ is only 1.8×10^{-6} at NTP, and for air $\mu'-1$ is 21 percent of this value. The values of $\epsilon'-1$ for O_2 and air, given in Table II, were decreased by the value of $\mu'-1$ given above to correct for the very small change in resonance frequency because of the magnetic susceptibility.

As a test of the internal consistency of the present measurements, the measured value of $\epsilon'-1$ of air, and the value computed from the measured value of the components of air are compared in Table III. It is seen here that the agreement between these values is well within the experimental errors. The values for $\epsilon'-1$ of N_2 and O_2 were obtained from Table II, and the value for A, whose contribution is small, was taken from the work of Jelatis.

C. Loss Factor of NH_3

The loss factor of NH_3 was measured at 9280 Mc as a function of pressure in the range 0 to 20 cm Hg. Results are shown in Fig. 5 for the two methods of measurement: (1) by measuring Q; and (2) by measuring the amplitude of the resonance curve. The data obtained by each method, which are shown in Table IV, agree within the errors (Sec. I.E4) for each method. It should be noted, however, that the two sets of data were taken at slightly different temperatures. It has been noted, in connection with the absorption of microwaves by O_2, that if the absorption is measured on the low frequency side of a group of lines, then ϵ'' should be related to temperature[24] by $\epsilon''\propto T^{-5/2}$. It is supposed that this relation applies to absorption by NH_3 at 9×10^9 Mc, which is an analogous case. Then it is found that for a given pressure ϵ'' at 23.5°C should be 1 percent higher than at 24.7°C. If this correction were applied to the data in Table IV, the results of the two methods would be in even closer agreement. It seems then, that the accuracy calculated for these methods in Part I may be on the pessimistic side.

[24] J. H. Van Vleck, Phys. Rev. 71, 413, reference 10 (1947).

TABLE IV. The dielectric loss in NH_3 by the Q and the amplitude method.

Pressure mm Hg	$\epsilon''\times10^4$ Q method $T=24.7$°C	$\epsilon''\times10^4$ Amplitude method $T=23.5$°C
200.1	0.967	1.015
180.1	0.799	0.841
160.5	0.648	0.668
140.5	0.482	0.518
120.7	0.393	0.375
80.9	0.177	0.167[a]
20.1	0.017	0.005

[a] For a pressure of 80.3 mm Hg.

The absorption in NH_3 at 9000 Mc is due to the contribution of many lines, the strongest of which are clustered about 24,000 Mc. This absorption can be represented by[25-27]

$$\alpha/\nu^2=\sum_n \beta_n p\{[\Delta\nu_n/((\nu_n-\nu)^2+\Delta\nu_n^2]+[\Delta\nu_n/((\nu_n+\nu)^2+\Delta\nu_n^2]\},$$

where $\alpha=2\pi\nu\epsilon''$ (cm^{-1}), β_n is an intensity factor (per cm Hg), p is the pressure (cm Hg), $\Delta\nu_n$ is the line width (cm^{-1}), and ν_n is the resonance frequency (cm^{-1}). At least for pressures up to 10 cm Hg, $\Delta\nu_n$ is proportional to pressure.[26] At these pressures and at a frequency removed from the resonance frequency, $\Delta\nu_n$ is negligible compared with the other terms in the denominators. Thus, the variation of α/ν^2 with pressure can be written as

$$\alpha/\nu^2=(\text{const.})p^2. \tag{10}$$

It was possible to represent accurately the data, given in Table IV, for pressures up to 10 cm Hg by such an equation in which the constant is 5.24×10^{-6} cm per (cm Hg)^2. The data used were obtained by the Q-method. Above 10 cm Hg, these data fall below the values given by Eq. (10).

ACKNOWLEDGMENTS

The authors wish to thank Mr. D. M. Kerns for helpful discussions on the theory of cavity resonators, also Mr. S. K. Chatterjee for assistance with the work on loss-factor measurements.

[25] J. H. Van Vleck and V. Weisskopf, Revs. Modern Phys. 17, 227 (1945).
[26] B. Bleaney and R. P. Penrose, Proc. Phys. Soc. (London) 59, 418 (1947).
[27] J. E. Walter and W. D. Hershberger, J. Appl. Phys. 17, 814 (1946).

Letters to the Editor

A Note on the Metallic Behavior of $MoSi_2$

FRANK W. GLASER
American Electro Metal Corporation, Yonkers, New York
November 8, 1950

DURING the course of an investigation conducted by this laboratory, electrical resistivity and the temperature coefficient of resistivity were measured for the compound molybdenum disilicide. This alloy was prepared by direct synthesis, i.e., Mo and Si powders of high purity were mixed in stoichiometric proportions of $MoSi_2$, cold pressed at approximately 3 psi and subsequently heated for 24 hr at 1300°C in a hydrogen atmosphere. This powder was then broken up and its average particle size brought to approximately 2 microns. Bars, approximately $1\times1\times2\frac{1}{2}$ cm, were then hot pressed, and a final density of 5.95 g/cc was measured on such specimens. The electrical resistivity was measured at various temperatures potentiometrically. The current employed during these measurements was of the order of 10 amp and the voltage drop was measured over a length of $1\frac{1}{2}$ cm. Table I shows linear dependence of the electrical resistivity on the temperature. The temperature coefficient is positive and the resistivity value of 21.5 microhm-cm at room temperature is comparable to that of a low carbon 1 percent Si-steel.

TABLE I. Electrical resistivity of $MoSi_2$ as a function of temperature.

Temp. °C	Electrical resistivity Microhm-cm	% increase per °C
−80	18.9	...
+22	21.5	0.135
+65	22.7	0.129

In view of the above, it must be concluded that not only interstitial compounds between transition metals and nonmetals have metallic properties (Hägg's rule). Compounds, not necessarily interstitial in structure—the radius ratio of Mo to Si being 0.86 as against 0.59 for compounds covered by Hägg's rule—can also have similar characteristics.

A Comparison of Quenching Distance and Inflammability Limit Data for Propane-Air Flames

DOROTHY MARTIN SIMON
National Advisory Committee for Aeronautics, Lewis Flight Propulsion Laboratory, Cleveland, Ohio
October 2, 1950

RECENTLY Friedman and Johnston[1] reported quenching distances of laminar propane flame as a function of pressure. They pointed out that the low pressure inflammability limits should correspond with their values if wall quenching determines the inflammability limit. In an obscure publication Huebner and Wolfhard[2] reported the low pressure limits of inflammability (propagation limit) for propane-air flames as a function of tube diameter. In Table I the low pressure limits of inflammability read from the composition-pressure curves for different tube diameters are compared with the extrapolated values of the minimum pressure for corresponding plane-parallel walls. Tube quenching diameters were assumed to be 35 percent greater than quenching distances for plane-parallel walls. The calculated values agree remarkably well with the experimental values for tube diameters from 33 to 89 mm; however, for smaller tubes the agreement is not so good, but the experimental values are quite uncertain for very small tubes. This correlation strongly suggests that wall quenching limits flame propagation in a tube.

TABLE I.

Tube diameter (mm)	Equivalent plane-parallel distance (mm)	Minimum pressure (mm) Experimental	Calculated
10	7.4	500–700	160
18.3	13.5	170–220	83
33	24.4	49	44
58	43.0	25	23
89	66	12	14

[1] R. Friedman and W. C. Johnston, J. Appl. Phys. **21**, 791 (1950).
[2] Huebner and Wolfhard, Volkenrode Monograph, Die Verbrennung in Grossen Holken, Issued by Technical Information Bureau for Chief Scientist, Ministry of Supply (1946).

Note on Static Electrification of Dust Particles on Dispersion into a Cloud

JOHN E. COOLIDGE AND GEORGE SCHULZ
The Pennsylvania State College, School of Engineering, State College, Pennsylvania
September 5, 1950

THE writers wish to call to Mr. Kunkel's[1] attention the work of V. E. Whitman[2] conducted in 1926, in which techniques similar to those described in the former's article were used to study the properties of the electrification of dust clouds. It is gratifying to see that, in a field in which contradictory results are common, consistent results were obtained after 24 years. To be sure, the measurements of Whitman were not as refined as Kunkel's so that quantitative measurements were not possible on individual particles, but the traces of the particles were obtained by time exposure. However, the general conclusion that the sign of the charge is independent of particle size but for dissimilar materials somewhat asymmetric in favor of the larger ones being negative is consistent with Kunkel's results.

There appears to be some contradiction in the effects of moisture content and dispersion methods on the resulting charge. For example, Whitman observed a large difference in net charge or asymmetry between wet and dry SiO_2. However, these differences could probably be resolved by a better understanding of the exact nature of the materials and methods used.

Another interesting comparison is the fact that both Whitman and Kunkel explained the charging action by purely statistical means. Here, Kunkel has extended the theory to a mathematical formulation but the general theories are the same. It might be noted that the reason offered for the lack of neutral particles in Kunkel's observations is obscure.

The writers wish to congratulate both Hansen[3] and Kunkel for the tremendous accomplishment in instrumentation described in their papers. It is felt that, when the tests are extended to include many more conditions, a great contribution to the understanding of frictional electrification will be made.

[1] W. B. Kunkel, J. Appl. Phys., **21**, 820 (1950).
[2] V. E. Whitman, Phys. Rev. **28**, 1287 (1926).
[3] W. B. Kunkel and J. W. Hansen, Rev. Sci. Instr. **21**, 308 (1950).

Comment on Static Electrification of Dust Particles on Dispersion into a Cloud

WULF B. KUNKEL
University of California, Berkeley, California
September 25, 1950

IN the preceding "Letter to the Editor" J. E. Coolidge and G. J. Schulz indicate certain points needing clarification in my recent article. I was aware of Whitman's work and regret I did not refer to it. However, since the sizes and charges of individual particles were not there determined, and owing to the

necessity of cutting down my paper to reasonable scope, it and the many other references to good but not directly relevant work were not included. The inclusion of such references would have entailed much discussion not cogent to the observations and their interpretation. The observations and interpretation were all that space would allow for in my paper.

In reference to the apparent contradiction between Whitman's and my work concerning the effect of moisture we should like to recall that Whitman did *not* determine the moisture content of his samples. He mentioned drying only, which we showed to be usually ineffective. As pointed out by Whitman himself, the difference in electrification noted by him for dried and undried quartz powder could as well have been caused by the difference in surface condition of the two brass tubes used in the experiment.

In order to understand why the simple theory of statistical random charging agrees with the observed scarcity of neutral particles we merely have to inspect the distribution equation

$$P(n)=\left(\frac{2}{\pi N}\right)^{\frac{1}{2}}\exp\left(-\frac{n^2}{2N}\right)=\alpha\sqrt{S}\exp\left(-\frac{n^2}{\pi\alpha^2 S}\right)$$
$$=\alpha(S)^{\frac{1}{2}}\left(1-\frac{n^2}{\pi\alpha^2 S}+\cdots\right),$$

in which we substituted $N=\frac{1}{2}\pi\alpha^2 S$ to express the proportionality between the number of contacts, N, and the surface area S. This means that unless $\alpha^2 S$ is of the order of n^2 or smaller, $n=0$ is not noticeably more likely than, say, $n=5$. Evidently $n=0$ is much less likely than $n\neq 0$, which proves the statement. Especially if $\alpha^2 S$ is relatively large, $P\simeq\alpha(S)^{\frac{1}{2}}$, i.e., independent of n for a considerable range of n, so that $|n|<10$ is not much more likely than, say, $500<n<520$.

A Condenser Aperture Device for Electron Microscopes

J. D. Boadway
Shawinigan Chemicals Limited, Shawinigan Falls, Quebec, Canada
October 26, 1950

THE main factor limiting the resolution of electron microscopes is considered to be spherical aberration of the objective lens. This may be minimized by use of a small angular aperture, either by using a fine diaphragm in the objective or condenser lens or by adjusting the illuminating electron beam so that the image of the source falls above or below the specimen.[1,2] The use of diaphragms has the difficulty that the fine apertures contaminate readily and that no adjustment is available in most modern instruments to align the hole with the optical axis of the lens. The adjustment of the condenser to move the crossover above or below the specimen has the difficulty that too large an area of the specimen is contaminated by electron bombardment.[3,4]

It has been observed that hot surfaces do not become contaminated by electron bombardment; hence, diaphragms which were heated would remain free from such contamination. The most convenient method of heating such diaphragms is by use of the energy in the electron beam. Although intensity of electron bombardment in the condenser lens, with a bias focused gun, is very high, the thick apertures normally employed carry the heat away too rapidly and become only warm; however, if the diaphragm used is sufficiently thin, this no longer becomes the case. A piece of 0.001-inch brass shim stock used as a diaphragm in the condenser lens rapidly lost its zinc by evaporation until a porous sponge of copper remained. A similar thin platinum diaphragm 0.001 inch in thickness remained free from contamination for over six months, in which time the hole became slightly larger and the crystals in the metal became large enough to be visible to the naked eye.

Such thin platinum could, hence, be used to reduce the angular aperture without suffering any contamination. However, unless the fine hole in this diaphragm were to coincide with the optical axis of the condenser lens, the illuminating electron beam would undergo a lateral motion as the condenser was focused and would be liable to shift with any instability in the condenser current. The special condenser aperture device shown in Fig. 1 makes such an adjustment possible in an RCA type EMU electron microscope, replacing the aperture and screen normally employed in this instrument.

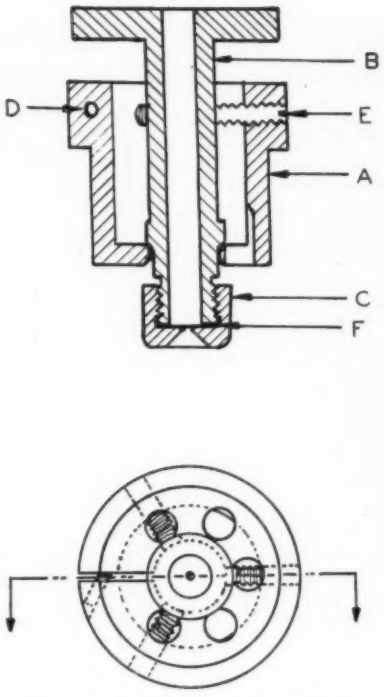


FIG. 1. Device for holding and permitting alignment of a condenser aperture in an electron microscope.

This device consists of an outer piece "A" which fits snugly inside the condenser and may be expanded to a tight fit by a screw "D", a central barrel "B", which is a working fit in the hole at the lower end of part A and is centered at the upper end by three screws "E", and a cap "C", which screws on the lower end of the barrel and holds the platinum diaphragm "F" in place. Part A holds the barrel to the geometric center of the lens at one point. The barrel may pivot about this point giving a lateral displacement to the aperture which it carries at its lower end. The three position screws E have a two-to-one leverage over motions of the aperture. Since it is found difficult to center the fine hole in the diaphragm over the hole in the cap, it is preferable to place the thin platinum disk in position and then to punch the hole with a fine pointed instrument. Indirect but large free air passage is provided through the device in order that the electron gun may be rapidly evacuated.

When this device is in the microscope, it may not be adjusted from outside; but this is not a serious disadvantage, since, once adjustments are made, they need not be changed for a year or so, that is, as long as the diaphragm remains in the instrument. Alignment of the aperture with the magnetic axis of the condenser may be made by adjusting the screws to move the image of the crossover in the same direction that it moves when the condenser current is increased. This requires that the direction in which the screws move this image be first determined and that ten to twelve adjustments be made to obtain an image of the source which does not move but merely expands and contracts as the condenser current is changed; for each adjustment it is necessary to strip down the top of the column to the condenser lens. As long as the hole in the diaphragm is not smaller than about 200 microns, the optical center of the magnetic lens will remain within the diaphragm hole or close enough to it that movement of the source is small.

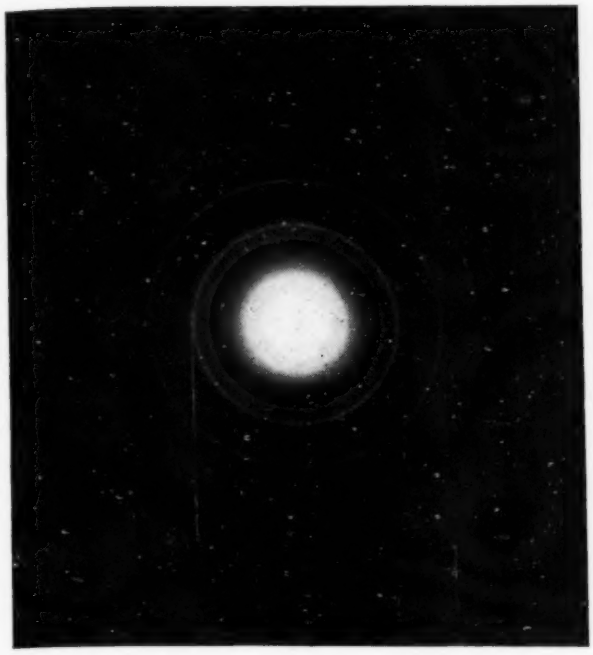


FIG. 2. Electron diffraction pattern of magnesium oxide obtained using a specimen in the object position of an RCA EMU microscope.

The use of a small, well-aligned condenser aperture has other advantages besides increased ease of operation of the instrument and better resolution. The optics of the illuminating condenser are so improved by this system that excellent electron diffraction patterns may be taken with the specimen in the normal position by removing the objective and projector pole pieces (Fig. 2). Since the beam may be focused to comparatively small area of the specimen, electron diffraction patterns may be taken of such areas or of single crystals (Fig. 3). If the electron gun and anode are clean so that the source is stable, it is possible to examine an area of a specimen and then to take an electron diffraction pattern of that same area.

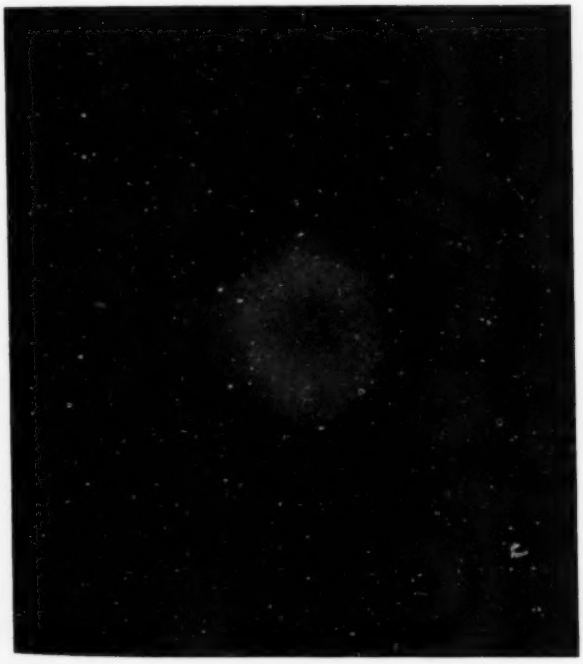


FIG. 3. Electron diffraction pattern of a small area of a specimen of powdered mica.

With a diaphragm in the objective and lateral displacement of the condenser lens, dark field images may be observed by the electron microscope.[5] These images are formed by scattered and diffracted rays rather than the electrons which pass directly through the specimen. The use of a finer condenser diaphragm leads to improved resolution in dark field and reduces the intrinsic line breadth of the diffracted cone, so that fewer bright crystal spots, crystals properly oriented to the beam, are observed and the amount of displacement of the condenser necessary to show the particular type of crystalline reflection becomes more critical. As the condenser is displaced from correct alignment, the alignment of the condenser diaphragm will no longer be correct unless by good fortune it is the correct height in the lens to lie on the optical center.

The use of a fine aperture made of thin platinum held in a centerable device thus reduces spherical aberration by use of smaller angular aperture without causing electron bombardment of large areas of the specimen. The thin platinum is held at a high temperature by electron bombardment so that it does not become contaminated. The optics of the condenser lens upon being improved makes possible better resolution in electron diffraction patterns produced with the specimen in the normal object position.

[1] J. Hillier and R. F. Baker, J. Appl. Phys. **16**, 469–483 (1945).
[2] J. Hillier and S. G. Ellis, J. Appl. Phys. **20**, 706 (1949).
[3] J. H. L. Watson, J. Appl. Phys. **18**, 153–161 (1947).
[4] J. Hillier, J. Appl. Phys. **19**, 226 (1948).
[5] C. E. Hall, J. Appl. Phys. **19**, 198 (1948).

Plasticity and Conductivity, Analogous Flow Phenomena in Copper Alloys

ROLLAND SYDNEY FRENCH
Research Department, Bridgeport Brass Company, Bridgeport, Connecticut
October 10, 1950

THE introduction of the solute atom to form a solid solution disturbs the periodicity of the solvent lattice and produces an increase in the electrical resistance.[1] Alloying also produces a lattice which exhibits a greater resistance to deformation and an increased ultimate strength. In a recent paper, French and Hibbard[2] studied the effect of various elements on the deformation

TABLE I. Chemical composition, strain hardening coefficient, and resistivity values of alloys.

Alloy	Percent solute Wgt.	Percent solute Atomic	Coefficient *m* soft, 0.090 mm	Resistivity microhm cm, soft	Source
Copper			0.575	1.7	a, b
Al	2.09	4.80	0.540	5.0	c
Al	5.96	13.00	0.510	11.3	a, b, d
Be	2.26	12.60	0.436	10.0	b
Cd	0.98	0.56	0.534	2.1	d
Ni	9.58	10.30	0.490	14.5	b
Ni	29.41	31.50	0.420	34.8	b
Ni	31.03	33.25	0.428	38.4	b
Si	1.14	2.55	0.550	8.0	c
Si	2.11	4.51	0.524	20.0	a
Si	2.97	6.47	0.514	24.8	a
Sn	1.02	0.74	0.520	3.6	a, b
Sn	5.37	3.95	0.448	10.0	a, b
Sn	8.44	6.27	0.436	13.7	a, b
Sn	9.90	7.38	0.430	15.3	a, b
Zn	9.95	9.69	0.512	3.95	a
Zn	20.60	20.15	0.490	5.41	a
Zn	30.68	30.05	0.482	6.4	a

[a] Copper and Brass Research Association Standards, Second Edition, 1949.
[b] Nat. Bur. Standards (U. S.) Circ. C 447 (1943).
[c] L. L. Wyman, Gen. Elec. Rev. **37**, 120 (1934).
[d] Bridgeport Brass Company.

of copper by means of the constants of the true stress—true strain equation:

$$\sigma = K\epsilon^m, \tag{1}$$

where σ and ϵ represent stress and strain, K is a constant, and m is the strain hardening coefficient. Since m defines by its own value the limiting strain of uniform deformation prior to necking, it has been looked upon as an index of plasticity. Data reported indicated that the plasticity of copper was decreased by the addition of a solute atom. Inspection has shown that these decreases were similar in form to the concomitant decrease in electrical conductivity. Should the foreign atom present an obstacle to the glide of the solvent lattice atoms in a manner similar to its effect upon the mean free path of the electron wave, a correlation of these properties might be expected. Table I lists the composition and determined m values which were shown in the recent paper, and electrical resistivity values.

As conductivity is the reciprocal of resistivity, the reciprocal of plasticity, or $1/m$, may be looked upon as a resistance to atomic flow. The data are shown in Fig. 1, plotted as the change in

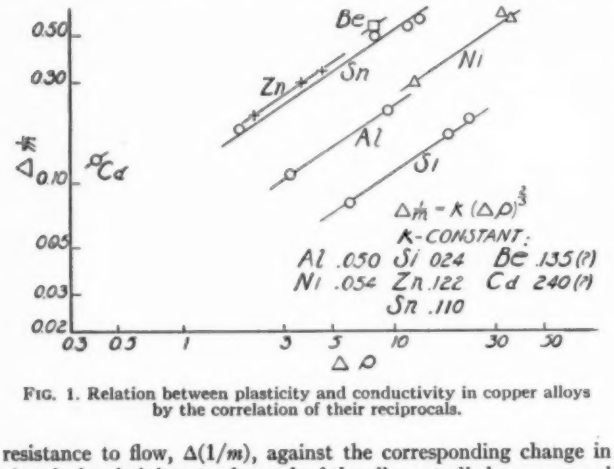


FIG. 1. Relation between plasticity and conductivity in copper alloys by the correlation of their reciprocals.

resistance to flow, $\Delta(1/m)$, against the corresponding change in electrical resistivity, $\Delta\rho$, for each of the alloys studied as compared to the values for copper. It will be seen that correlation is evident of the form:

$$\Delta(1/m) = k(\Delta\rho)^1. \tag{2}$$

It is surmised that since plastic flow essentially involves three dimensions of the specimen, whereas conductivity is a function of two, the solute atom's effect necessitates correlation based upon these parameters and Eq. (2) could be written more accurately:

$$[\Delta(1/m)]^3 = \kappa(\Delta\rho)^2. \tag{3}$$

It would seem, therefore, that the dispersion of solute atoms in copper controls the atomic flow of uniform deformation in an analogous manner to its control of the mean free path of the electron. As the concentration increases, both plasticity and conductivity are decreased accordantly.

[1] W. Hume-Rothery, Institute of Metals, Monograph and Report Series No. 3, 204 (1946).
[2] R. S. French and W. R. Hibbard, Jr., Trans. A.I.M.E. 188, 53 (1950).

A Noncontact Temperature Measuring Device

J. E. FIELD AND S. D. GEHMAN
Research Laboratory, Goodyear Tire and Rubber Company, Akron, Ohio
October 10, 1950

THERE are numerous occasions in industrial processing when it is required to measure, record, or control the surface temperatures of moving rolls, belts, or materials. Any method of temperature measurement involving contact with such moving surfaces is subject to obvious inaccuracies and practical difficulties. The temperatures are often too low for the application of conventional methods of radiation pyrometry. A low temperature radiation pyrometer is a rather elaborate and expensive instrument.[1]

The following relatively simple type of radiation thermocouple receiver has been found to be useful for such applications, especially where temperature measurements are required at a large number of stations. It differs in principle from a radiation pyrometer in that the receiver is mounted close enough to the moving surface to approach temperature equilibrium with it. By this expedient, the sensitivity of the receiver is increased as compared to that of a pyrometer because the receiving element undergoes approximately the same temperature changes as those of the surface being measured. In an ideal case of the application of this principle, the equilibrium response of the receiver would be independent of the emissivity of the surface being measured. Actually, a small dependence on surface characteristics is observed. This becomes less important with each feature in the design which reduces the temperature difference between the receiver and the surface and which makes radiation between them the dominant mode of heat transfer for the receiver element. .

Figure 1 is a diagram which illustrates the principles of construction of the device. Ambient temperatures may affect the readings of such a receiver due to (1) radiation transfer to the receiver from any surroundings other than the surface to be measured, (2) air currents, (3) heat conduction through the mounting or the thermocouple wires.

The receiving elements are thin copper foils, about 1 mil thick, blackened on the front surface and brightly plated on the rear surface. In use, they are spaced about 3/16 in. from the surface under observation. These features minimize radiation transfer to the receiver from any surface except that for which the temperature is to be measured.

The receiver may be protected from air currents by a fabric skirt around the mounting which is nearly in contact, or may actually make light contact, with the moving surface.

Heat conduction to the receiver is reduced by using fine (No. 40) thermocouple wire and by mounting the receiver about 1/8 in. above a Celotex disk by means of very light wooden pegs or by means of the thermocouple wires themselves.

In spite of these precautions, there is a residual effect due to

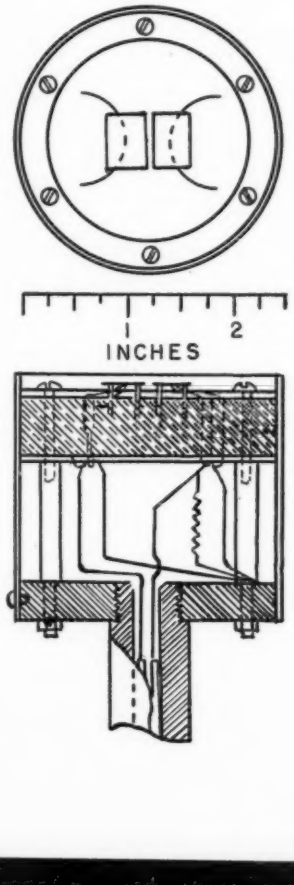


FIG. 1. Views of radiation receiver.

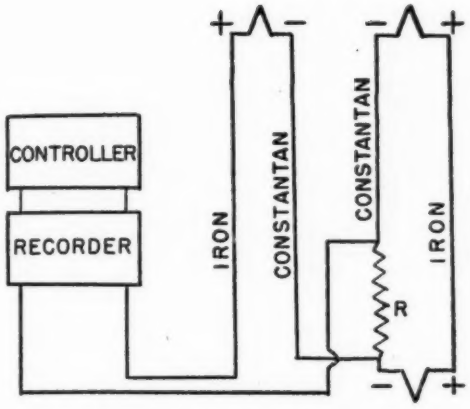


FIG. 2. Circuit diagram for thermocouples.

ambient temperature. To illustrate the magnitude of the effects, a reading of 175°F was obtained when the housing temperature was 210°F and the true surface temperature was 165°F. The necessary corrections could be accomplished by providing a thermostated housing and calibration curve, but a simple compensating circuit was found to be adequate as long as the ambient temperature did not fluctuate too rapidly. This method has the advantage of permitting the use of the device with conventional potentiometer type recorders and controllers.

The circuit diagram in Figure 2 explains how compensation is effected. Assuming that the ambient temperature is higher than the surface temperature to be measured, it is evident that the temperature indicated by thermocouple 1 will be higher than the true surface temperature. In order to record the correct temperature, the voltage generated by thermocouple 1 must be reduced to the proper value at the terminals of the recording instrument.

This is accomplished by introducing a compensating receiver 2, which is identical with receiver 1 in every respect. The compensating receiver is connected in a circuit which includes couple 3, which is exposed to the ambient temperature. When the ambient temperature differs from the receiver temperature, a current will flow in the compensating circuit. The magnitude of this current is proportional to the temperature difference. The constantan resistance R is inserted in the compensating circuit. The measuring thermocouple 1 is connected in series with this resistance, the polarity being arranged so that the voltage drop across R is subtracted from the voltage generated by thermocouple 1 when couple 3 is at a higher temperature than couple 2. The correct value of R can be calculated from the circuit resistance and the temperature correction involved. Thus in the previous case a 10°F correction was required when the temperature difference in the compensating circuit was 210 to 175 or 35°F. Therefore, R should be the fraction 10/35 of the total resistance of the compensating circuit. In the units which have been constructed, R has been about 10 ohms, the final adjustment being made in a calibration set-up using a stationary surface at a controlled temperature, the ambient temperature also being controlled. The compensation will then apply accurately and automatically for a wide range of temperatures.

The output voltage of the device can be impressed across the terminals of a potentiometer or noncurrent drawing type of recorder or controller to read directly the correct temperature. A number of these simple receivers have been constructed and used to record and control the temperature of a moving belt within ±2°F in a range of 140 to 180°F, the ambient temperature being about 250°F.

The speed of response depends upon the heat capacity and area of the receiver. With the construction described, 80 percent of full response was attained in 5 sec.

[1] J. C. Mouzon and C. A. Dyer, J. Opt. Soc. Am. 39, 203 (1949).

A Magnetic Scaling Circuit

HELLMUTH HERTZ
Department of Physics, University of Lund, Lund, Sweden
November 8, 1950

FOR the counting of random pulses at high rate from a G-M tube different kinds of scaling circuits are used between the G-M tube circuit and the mechanical counter. The storing of pulses can be done in different ways, for example, in electric tube circuits, in condensers,[1] or in ferromagnetic materials.[2] Using the last method a scaling circuit has been developed on the following basis.

If a condenser C is discharged through a gas-filled rectifier tube and the primary coil P (350 turns) on a small ring-formed iron core (Fig. 1), a voltage pulse is induced in the secondary coil S (500 turns). Its time-dependence as measured between A and B is shown in Fig. 2, curve 1. If the condenser used is not too large, the iron will become magnetized by the discharge only to a certain degree and not to its saturation value. If the condenser is now discharged a second time, a new voltage pulse will be induced in S as shown by Fig. 2, curve 2. The result of a third condenser discharge through P is shown by curve 3, of a fourth

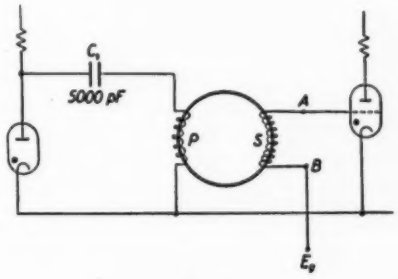


FIG. 1.

by curve 4, etc. As is seen, the maximum voltage V_n of the nth pulse increases with each following pulse and approaches asymptotically a limiting value V_∞ as the iron becomes fully saturated. The form of these curves depends, of course, on the parts used in the primary circuit, for example, the iron core, the number of turns on the coils, etc. If the coil S is connected to the grid of a thyratron tube in such a way that the voltage V_n shown in Fig. 2 drives the grid positive, the tube will fire after the first, second $\cdots n$th pulse depending on the value of the negative voltage E_g

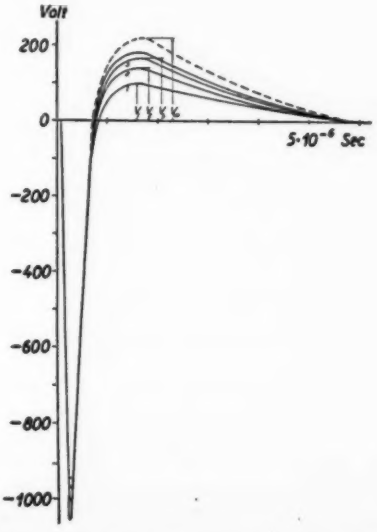


FIG. 2. Induced voltage impulse measured between A and B in Fig. 1.

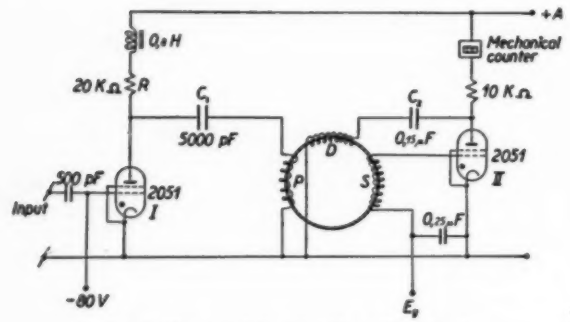


FIG. 3. The complete magnetic scaling circuit.

on the grid of the thyratron tube II. This effect works even better if one starts at negative saturation of the iron and drives it to positive saturation by the condenser discharges.

A scaling circuit using this effect is shown in Fig. 3. Every time a short positive pulse is applied at the input circuit the thyratron I fires and the condenser C_1 is discharged through the coil P, thereby partly magnetizing the iron core which at the beginning was at negative saturation. Since the discharge circuit is a LC circuit, the thyratron discharge is quenched automatically when the anode becomes negative. Thus, the recharging resistance can be held at a relatively low value. At each discharge of the condenser C_1 a positive voltage V_n is induced in S, which would fire the thyratron II, if the negative voltage E_g did not prevent it as long as $V_n+E_g<E_f$, where E_f is the ignition voltage of the thyratron II. Suppose now, a scale ratio of N:1 is desired. Then the value of E_g is chosen so that $E_g+V_n\geq E_f$ for $n\geq N$, but for all $n<N$, $E_g+V_n<E_f$. Thus, first the Nth pulse will fire the thyratron II, thereby discharging the condenser C_2 through the demagnetization coil D (1000 turns). By choosing the condenser C_2 much larger than C_1, it is possible to drive the iron core to negative saturation in about 3.10^{-6} second by this discharge. That brings the iron core to the starting point again and the next N pulses can be counted. By the recharging circuit of the condensor C_2 a second scaler stage can be triggered or a mechanical counter directly operated. By using small thyratron tubes the scaler can be built into a small unit.

Scale ratios up to 25:1 have been obtained with one scaler stage, but then the ratio stability was poor. Probably no higher ratios than about 10:1 are stable enough for practical use, while ratios of about 5:1 were absolutely stable. Pulses following each other in 2.10^{-4} second can be separately counted. Difficulties encountered were exclusively due to instabilities of the thyratrons and not to properties of the iron core. In the present apparatus these difficulties are probably accentuated by the fact that that the peak current through the thyratron II at the discharge of the condenser C_2 is higher than permissible. Thus, even though the condenser discharge is of very short duration, counting rates of more than 2000 pulses per second over long periods will damage the tube. By changing the dimensions of the iron core it will probably be possible to eliminate this difficulty.

A quantitative explanation of the underlying magnetic processes has been given.

A more detailed account of this work will be published in the *Arkiv för Fysik*.

[1] Lewis and Raffle, J. Sci. Instr. **27**, 7 (1950).
[2] An Wang and Day Dong Woo, J. Appl. Phys. **21**, 49 (1950).

Erratum: Theory of the Parallel Plane Diode

[J. Appl. Phys. **21**, 974 (1950)]
A. H. TAUB AND NELSON WAX
University of Illinois, Urban, Illinois

THE conditions for a single valued velocity theory given in the above paper were stated, incorrectly, to be both necessary and sufficient. The condition [inequality (4.7)] and its consequences [inequalities (4.10) and (4.11)] are necessary and sufficient conditions for a single valued velocity theory over the entire $x-t$ plane, and are only sufficient conditions in the region between cathode and plate. Mr. Alan D. Sutherland has kindly called our attention to the error in respect to the space charge limited diode. The necessary and sufficient condition, for all cases, is given below.

Using our previous notation, let $\bar{\xi}$ be the values of ξ satisfying $\partial x/\partial\xi=0$, namely, the real roots of

$$\tfrac{1}{2}(t-\bar{\xi})^2[h(\bar{\xi})-dE_0(\xi)/d\xi|_{\bar{\xi}}]-(t-\bar{\xi})[\tfrac{1}{6}\kappa^3 dv_0(\xi)/d\xi|_{\bar{\xi}} -E_0(\bar{\xi})]+\tfrac{1}{6}\kappa^3 v_0(\bar{\xi})=0. \quad (1)$$

The envelope of orbits, $X(t)$, will then be given by

$$\tfrac{1}{6}\kappa^3 X(t)=\int_{\bar{\xi}}^{t}\tfrac{1}{2}(t-t')^2h(t')dt'+\tfrac{1}{2}(t-\bar{\xi})^2E_0(\bar{\xi})+\tfrac{1}{6}\kappa^3(t-\bar{\xi})v_0(\bar{\xi}), \quad (2)$$

with $t>\bar{\xi}$.

Let $\bar{t}$ be the times at which some orbits intersect. These orbits can be identified by the real roots, ξ^i, of

$$X(t)=x(\xi^i,t). \quad (3)$$

The inequalities

$$0<x(\xi^i,t)<1 \quad (4)$$

for all $t\leq\bar{t}$ require that electron trajectories remain in the cathode anode region before, and at the times, they intersect.

Similarly, the inequalities

$$0<X(t)<1 \quad (5)$$

also require points on the envelope to be in the cathode anode region.

Therefore, the necessary and sufficient condition that any two electron orbits shall not have a point in common in the space between cathode and anode is that no points $[t, X(t)]$ can be found which satisfy inequalities (4) and (5).

Spectral Emissivity of Coatings of Thoria and Other Refractories as a Function of Temperature

F. H. MORGAN
Bartol Research Foundation of the Franklin Institute, Swarthmore, Pennsylvania
July 17, 1950

MEASUREMENTS are here reported for the spectral emissivity of coatings of several refractory compounds on tungsten or molybdenum filaments, in vacuum, as a function of temperature. It seems noteworthy that, unlike the other compounds studied, the thoria coating shows a decided increase of emissivity with temperature. Studies of the relation between thermal and thermionic properties of thoria are being reported elsewhere.[1]

A prior study of the emissivity of bulk thoria[2] was performed in air and showed different temperature dependence than the coatings discussed here.

Curves of brightness *vs* true temperature were obtained by two methods. The first employed tungsten-tantalum and tungsten-moly thermocouples,[3] prepared from 2-mil or 5-mil stock, spot-welded to a 30-mil tungsten filament. The filament was then coated with the material under consideration, mounted in a bell jar or tube and evacuated. Pyrometer measurements of brightness temperature were taken at the point of the thermocouple-filament weld.

In the second method[4] a long directly heated tantalum sleeve, with a narrow slit in it, was coated. Radiation from the interior of the cylinder was assumed to approximate blackbody radiation and pyrometer readings through the slit were accepted as true temperature. This method was checked by spot-welding a thermocouple to the outside surface of the cylinder, and found to be satisfactory. From these curves the spectral emissivity (0.655μ) can be easily calculated using the usual relation.

Table I gives values of spectral emissivity for several materials

TABLE I.

Material	Spectral emissivity at 0.655 micron
Thoria	0.40
Zirconium carbide	0.96
Titanium carbide	0.96
Tantalum carbide	0.67
Tantalum boride	0.70
Zirconium boride	0.70

For all materials except thoria, a single value of emissivity is obtained at all temperatures in the experimental range. Figure 1 is a typical curve for zirconium boride and tantalum boride.

In the case of thoria, 2-mil tungsten and moly thermocouple wires were spot-welded to a 20-mil tungsten filament which was flashed in hydrogen and then coated cataphoretically with thoria.

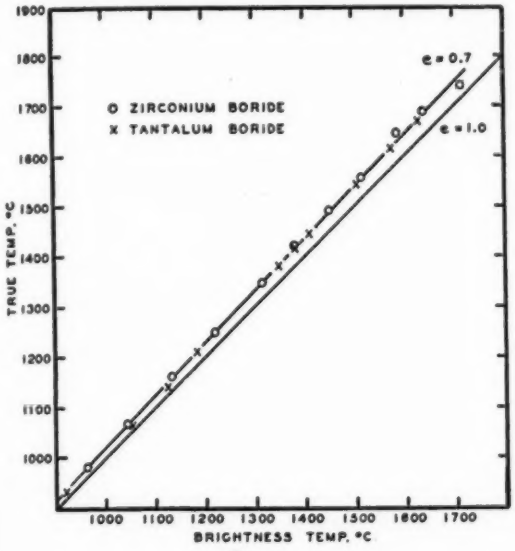


FIG. 1.

This was then evacuated on a mercury system to avoid contamination by oil vapors. Figure 2 shows, for thoria, the temperature dependence of the spectral emissivity. The emissivity (ϵ) was found to be 0.4 up to 1800°C at which point an increase in temperature brings about an increase in spectral emissivity. At 2040°C, $\epsilon = 0.62$ and is still increasing with further increase in temperature. If the filament is not subjected to temperatures higher than cited above, the curve may be run repeatedly and the data are reproducible, provided that the filament has been thoroughly outgassed. (Weinreich[1] has found that flashing to 2600°C results in a higher emissivity which persists at the lower temperatures.) Different lots of thoria have been found to give essentially the same pattern of increasing emissivity with temperature although the absolute value at 1600°C (for example) may vary between 0.3 and 0.45.

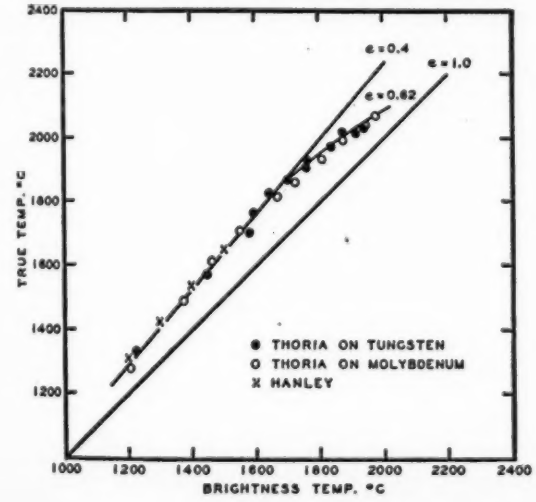


FIG. 2.

The experiment was repeated using a molybdenum filament coated with thoria, giving identical results. These data are represented in Fig. 2. Hanley[5] has also investigated the spectral emissivity of thoria up to 1600°C, and some points from his data are recorded in Fig. 2. He has reported a value of 0.35 for the spectral emissivity.

[1] O. A. Weinreich, J. Appl. Phys. (to be published).
[2] G. Liebman, Z. Physik 71, 416 (1931).
[3] F. H. Morgan and W. E. Danforth, J. Appl. Phys. 21, 112 (1950).
[4] A. G. Worthing, *Temperature* (Rheinhold Publishing Corporation, New York, 1941), p. 1164.
[5] T. E. Hanley, J. Appl. Phys. 19, 583 (1948).

The Natural Width of a Grain Boundary in a Metal

J. L. SNOEK
Department of Physics, Horizons Inc., Cleveland, Ohio
October 10, 1950

THE natural width of a grain boundary in a metal is defined here as the width observed in the complete absence of internal strains and of impurities.[1] This width may be expected to be a function of the relative orientation of the adjacent crystals only, the boundary itself having ample opportunity during annealing to choose the orientation leading to minimum surface energy. Our considerations are restricted to small angles.

In the analogous problem of determining the width of a boundary zone between magnetic domains—the so-called Bloch wall—the width which establishes itself is the result of two opposing tendencies.[2]

The problem is to find those opposing tendencies in the structure of the grain boundary. It is believed that these opposing tendencies have been found.

Cahn's experiments on polygonization[3] show that uniform arrays of parallel edge dislocations on annealing tend to agglomerate and thus form crystal boundaries of practically monatomic width.

The counterpart experiment with screw dislocations has recently been carried out by D. C. Jillson.[4] Whereas Cahn bent his zinc crystals around an axis in the basal plane, thus producing edge dislocations, Jillson twisted zinc crystals around an axis normal to this plane and thus produced a uniform array of screw dislocations.

According to Jillson, there is no tendency to form any discontinuity under these conditions. This indicates that screw dislocations of equal sign contrary to edge dislocations tend to repel each other. The grain boundary thus tends to occupy the whole of the length of the crystal, and the natural width under these ideal circumstances should be infinite.

The width observed in the general case in which edge dislocations as well as screw dislocations are present is probably determined by the two opposing tendencies indicated above.

Experiments intended to verify our fundamental conjecture, namely, that no polygonization exists for zinc crystals which are twisted around an axis normal to the basal plane, are being started.

To avoid misunderstanding, it should be added that is is known that parallel dislocations of equal sign in infinite crystals always repel each other, but that the mathematical problem of determining the interaction in crystals of finite dimensions is unsolved.[5]

[1] See the discussion between K. W. Andrews and A. H. Cottrell in Report of a Conference on the Strength of Solids, Bristol, 1948, p. 38.
[2] See R. Becker and W. Doring, *Ferromagnetismus* (Berlin, 1939).
[3] R. W. Cahn, Bristol Report, p. 136.
[4] D. C. Jillson, J. Metals 188, 1005 (1950).
[5] See A. H. Cottrell's article in Bruce Chalmers, *Progress in Metal Physics*, Vol. 1 (London, 1949).

PROGRAM

Proceedings of the Electron Microscope Society of America

THE annual meeting of the Electron Microscope Society of America was held at the Statler Hotel in Detroit, Michigan, September 14, 15, and 16, 1950. Titles and abstracts of the papers presented are given below. The name of the person presenting the paper is marked with an asterisk.

Electron Metallography and Related Subjects

CHAIRMAN: CHARLES H. GEROULD

1. Some Factors Affecting the Study of Martensite and Austenite. A. L. ELLIS* AND L. E. SMITH, *Manufacturing Research, International Harvester Company.*—This study is concerned with some of the effects of tempering on the differential etching of martensite and austenite. Specimens of two different compositions of steel were quenched and then tempered. Micrographs are presented which show the difference in structural detail, which resulted from tempering variations. Evidence presented indicates that the degree of differentiation of martensite from austenite is dependent upon tempering for the electron microscope as well as the light microscope. The effects of etching and cleaning are discussed.

2. A Method of Lineal Analysis with the Electron Microscope. F. K. IVERSON* AND A. L. ELLIS, *Manufacturing Research, International Harvester Company.*—A summary of the principles of lineal analysis and some of the special problems which arise in applying it to the electron microscope are discussed. A description of the Hurlbut counter and its use with the electron microscope is presented. The experimental procedure used to determine volume relationships of phases in steel is described. The reproducibility of the lineal method is discussed.

3. The Electron Microstructure of an Eutectoid Steel. GEORGE E. PELLISSIER (For joint Subcommittee XI of ASTM Committee E-4), *Carnegie-Illinois Steel Corporation, Pittsburgh, Pennsylvania.*—For the past two years, several research laboratories have cooperated, as a joint committee (now Subcommittee XI of ASTM Committee E-4 on Metallography), to investigate the potentialities of the electron microscope as a means of obtaining new information on the microstructure of steel. The fidelity and reproducibility of existing methods of sample and replica preparation have been evaluated by comparing electron micrographs made by these methods, with each other, and with the best obtainable light micrographs on identical structures; where necessary, these methods were modified or improved. Those methods found to be most reliable and powerful were used to investigate the fine-scale morphology of the decomposition products of austenite and martensite in a plain carbon eutectoid steel. Electron micrographs, at a magnification of 15,000 diam of fine pearlite, bainite, and tempered martensite, clearly show the size, shape, and arrangement of carbide particles in the ferrite matrix.

4. Selection of Etchants for Electron Metallography. WILLIAM L. GRUBE, *Research Laboratories Division, General Motors Corporation, Detroit, Michigan.*—An important factor in electron metallography is the proper selection of etchants to develop the microstructure of the specimen. In contrast to light metallography, no characteristic films or stains can be tolerated on etched specimens to be replicated because of the danger of introducing artifacts and the difficulties of removing replicas. Therefore, the etchant must be capable of developing a relief upon the surface which, when reproduced as minute differences in thickness of the replica, is sufficient to differentiate between the microconstituents. When only two microconstituents are present, the choice of a satisfactory etchant is not difficult. When three or more are present, the proper choice becomes more involved. The nature and importance of the problem are discussed as well as the characteristic action of various useful etchants upon particular microconstituents. The comparison of etchant characteristics is illustrated with electron micrographs of representative ferrous and nonferrous metal specimens.

5. A Technique for the Study of Etched Cellulose Acetate Fibers. WILBUR KAYE* AND VIRGIL PECK, *Research Laboratories, Tennessee Eastman Corporation, Kingsport, Tennessee.*—A technique is presented for the preparation of high resolution replicas of the surfaces of etched cellulose acetate fibers. Etching is achieved by immersing the fibers in acetone cooled to $-35°$ to $-40°$C for a few minutes. The etching process is stopped by flooding with cold ethyl alcohol. Replicas of these surfaces are prepared directly by the deposition of a uranium shadow layer backed with an aluminum-beryllium alloy. The fiber is subsequently dissolved with cold acetone. Slides are presented showing the dispersion of pigment in the yarn and possible crystalline areas.

6. The Preparation of Substrates and Replicas for High Resolution Studies. J. J. COMER* AND F. A. HAMM, *General Aniline and Film Corporation, Easton, Pennsylvania.*—A method is described for preparing silicon monoxide and silicon dioxide films by the evaporation of these materials onto the surface of freshly cleaved mica. These films of various thicknesses and shadowed with equal amounts of Pt-Pd have been compared from the standpoint of inherent structure. The results have indicated that a relatively thick (200A) silicon dioxide film exhibits the least structure and may be useful in high resolution studies of certain biological specimens or in size determination of small particles of relatively high scattering power. Thin silicon monoxide films (40A) were found to exhibit more structure than the thick silicon dioxide films, but less than the usual collodion films. For this reason they are desirable for use as a substrate with substances of low electron scattering power when greater contrast is needed. The need for careful control of the thickness of the shadowing metal is illustrated by micrographs showing how background structure may be introduced by the metal itself as the thickness of the evaporated metal is increased.

7. An Experimental Study of Image Contrast at 50, 100, and 150 kv. JAMES HILLIER AND S. G. ELLIS*, *RCA Laboratories, Princeton, New Jersey.*—Step-films of silicon monoxide were prepared by evaporation. Their thicknesses were determined by a Tolansky multiple beam interference method. The densities of similar films were found by weighing. From electron micrographs at 50, 100, and 150 kv, the image current density conjugate to the steps and to clear space were compared. The results show that it is advantageous to use the lowest voltage in the study of thin specimens. The increase in

penetration at the higher voltages is offset, in part, by the lowering of the contrast. The information from thicker specimens may also be limited by overlapping of the structures examined. The work has revealed some technical advantages in using higher voltages, such as decreased specimen loading, some reduction in sensitivity of the beam to disturbing factors and, within limits, increased photographic sensitivity. A complete analysis cannot be made until the variation of resolution with accelerating voltage has been studied experimentally.

Instrumentation

CHAIRMAN: G. D. SCOTT

8. A Small Electron Microscope. JOHN H. REISNER* AND EDMUND G. DORNFELD, *Radio Corporation of America, Camden, New Jersey.*—A new electron microscope of greatly simplified design has been constructed. The instrument has a resolution of 100A, yet is small enough to be mounted on a desk or laboratory table. Magnetic lenses are energized by permanent magnets. An accelerating voltage of 50 kv is varied over a narrow range to provide means of focus. Images may be photographed on 2×2 plates. Plates may be changed in a minute, and specimens in fifteen seconds. Direct magnifications of 1500, 3000, and 6000 are possible depending on pole pieces used. Photographic images may be enlarged up to ten times. The instrument is free of *x*-radiation and completely interlocked and provided with shorting devices for protection from high voltage.

9. A New Type of Shadow Microscopy. M. J. COLUMBE, *General Engineering and Consulting Laboratory, General Electric Company.*—A simple method for viewing samples in an electron diffraction instrument is described. Magnification of the sample up to several hundred diameters has been obtained by placing a thin film between the electron source and the sample. The sample may then be evaluated in terms of electron transparency, particle dispersion, and crystal habit to locate the best area for taking diffraction patterns. Illustrations will be shown.

10. An Electron Microscope Photometer. F. A. HAMM, *General Aniline and Film Corporation, Central Research Laboratory, Easton, Pennsylvania.*—This paper describes an electronic device for measuring the intensity of the focused electron beam as it passes through the final viewing chamber of the RCA electron microscope Model EMU. A portion of the electron rays serving to form the final image is collected on a metal probe; the electrons are passed to ground through a high resistor. The voltage developed across this dropping resistor is measured and is a direct measure of the intensity of the focused electron current. The uses for this photometer are as follows: (1) as an exposure meter in electron microscopy photography; (2) as a means for determining the relative or absolute thicknesses of electron microscope specimens, (3) as a means for evaluating the energy incident to the electron microscope specimen so that it may be possible to correlate changes (polymorphic transformations, polymerizations, contaminations, etc.) with this incident energy.

11. A New 100-Kilovolt Electron Microscope. J. H. REISNER,* R. G. PICARD, AND E. G. DORNFELD, *Radio Corporation of America, Camden, New Jersey.*—A high resolution electron microscope employing high and medium accelerating voltages and providing a maximum of flexibility along with constructional and operational simplicity is described. Provisions for large photographic plates and an extended stage motion permit large fields to be observed and photographed. A built-in magnification viewer permits accurate focus of images to maximum resolution from observation of contours. Extended range lenses permit direct magnifications from 1000 to 20,000 times at 100 kv. A two-stage gun and double element condenser lens system permit great flexibility in beam control. Power supplies and control circuits are located remotely from the column for the reduction of fields and temperature rises. Disassembly of the vertically disposed column for routine maintenance and cleaning can be done rapidly without tools. Simple, rigid locking controls permit perfect realignment following disassembly. Direct viewing of the image on a 4×5 screen permits the brightest possible image for a given beam intensity. Cassette holds four 2×2 or 3¼×4 plates permitting a permanent record on a dimensionally constant emulsion. Plates are exposed at viewing screen magnifications to keep subsequent optical magnification to a minimum. The vacuum system includes an auxiliary water-cooled trap to minimize back diffusion of vapors from the pumping systems. Magnetic vacuum valving is provided.

12. A 100-kv Electron Microscope Voltage Supply. S. M. ZOLLERS,* H. E. REEBER, AND H. F. SCHNEIDER, *Radio Corporation of America, Camden, New Jersey.*—A dc high voltage supply has been constructed as a driving source of voltage for a new electron microscope. The supply has a stability to 1 part in 50,000 and can operate at 50 or 100 kv at a load of half a milliampere. A unique geometric arrangement of rectifiers and capacitors limits the potential difference across dielectric surfaces to 50 kv permitting reliable use of air insulation in the supply. High rf voltage is generated by applying a 40 kc signal across a high *Q* inductor, and rectifying the voltage in a four-tube cascade. Filament power of three amperes is supplied by a 200 kc oscillator. Rectifier filaments are heated from the rf space currents in the supply. A resonance filter and a high gain feed-back amplifier provide a stable output voltage regardless of load or line fluctuations. All stabilization reference is against batteries for the ultimate in stability. Simplicity of structure and use of standard components permits quick and easy servicing of the supply.

13. Rapid Non-Photographic Measurement of Specimen Dimensions. HENRY FROULA, *Department of Engineering, University of California, Los Angeles, California.*—This paper discusses methods for the rapid measurement of specimen dimensions directly on the fluorescent screen of the electron microscope; i.e., without the use of photography. In particular, the techniques and results of two such methods are described: In the first of these the specimen dimension to be measured is magnified to a given size by means of a continuously variable magnification control. In the second method, the image size remains fixed while, by means of an external control, a 0.200-in. circular aperture is moved along the axis between the projector lens and the final screen; at some unique position of this apertured diaphragm its shadow exactly circumscribes the image dimension under investigation. In either method the final position of the external control can readily be interpreted in terms of the specimen dimensions.

14. A Metal Evaporator Using High Frequency Induction Heating. R. G. PICARD* AND J. E. JOY, *Radio Corporation of America, Camden, New Jersey.*—The difficulty of evaporating platinum from heated filaments and the desirability of trying this metal for shadow casting led to the development of a metal evaporator using high frequency induction heating for the charge. An rf generator was used to supply power to the metal through a special coupling coil. A new type feed through bushing for getting the rf power into the vacuum enclosure was designed. Successful evaporations of small charges of metal were made. The deposit is free of contaminating materials which occur during evaporation from heated filaments or boats.

15. Improving Precision of Powder Diffraction with the RCA-EMU Electron Microscope. M. S. JAFFE, *Lamp Department, General Electric Company, Nela Park, Cleveland, Ohio.*—Precision of about 0.1 percent in measuring diffraction rings can be attained with the electron microscope if certain precautions are taken. The particle size of the specimen and the size of the source must be optimal to produce fine rings. The measuring instrument must be accurate to 0.01 mm. The electrical circuits must be carefully warmed up to avoid drift in ring diameter. The reference standard must be in the plane of the specimen, so that the standard pattern can be taken before and after the unknown without changing any electrical adjustments. Stray magnetic fields must be shielded or compensated to avoid elliptical rings. A graphical method must be used to compensate for the pincushion or barrel distortion introduced by the lenses.

16. Effect of Coil Current Reversal on Distortion in an EMU-2A Electron Microscope with Extended Range Lens. J. R. COOPER, *Lamp Department, General Electric Company, Nela Park, Cleveland, Ohio.*—The eight possible combinations of field direction were studied for their effect on distortion. It was found that the distortion changed in magnitude and in the relative positioning of the center of distortion in reference to the alignment center. One arrangement of polarities was found to give both least distortion and best coincidence of distortion center with the field center. For this instrument this agreement occurred when the condenser lens field was South up, the objective South up, the extended range lens North up, and the projector lens South up. The results obtained indicate that the extended range lens will give a low power image with minimal distortion only when it is used with the most advantageous polarity combination. With the least desirable arrangement the percent distortion is increased three to four times.

17. Latex Size by X-Rays. K. L. YUDOWITCH, *Florida State University, Tallahassee, Florida.*—Latex spheres in common use as magnification standards have been measured by light and electron microscope, light scattering and ultracentrifuge. The reported diameters range from 2520 to 3210A with probable errors of 1 to 7½ percent; the accepted average is 2590A. The spheres have been measured by x-ray scattering. From the positions of the scattered rings, the computed average diameter is 2780A ±0.06 percent. This value is obtained with a small angle x-ray scattering camera of high resolution, and the results corrected for collimation error. The increase of this value over the accepted electron microscope value of 2590A gives possible support to the contention that the electron microscope results suffer from an electrostatic artifact.

18. Microscopical Resolution of the Morphology of the Test-Diatom, *Pleurosigma Angulatum*. T. G. ROCHOW,* A. F. KIRKPATRICK AND F. G. ROWE, *Research Laboratories, American Cyanamid Company, Stamford, Connecticut.*—The diatom, *Pleurosigma angulatum*, has long been used to test the resolving power of the microscope using **transmitted light.** The units of periodic structure have been described as hexagonal, circular or elliptical in cross section, depending on the optical conditions of illumination and observation. By **reflected light,** however, the units appeared to be small or large ellipses, or slits, depending chiefly on the level of focus. **Electron micrographs** of **replicas** of the half-shell showed that on one side there were tiny elliptical openings and on the other side there were slit-openings. **Electron micrographs** of **fragments** of the shells showed that the unit-chamber was elliptical in most planes parallel to the shell. A model will be shown to summarize the evidence. Practically all the appearances by reflected light were explained by electron micrographs. Most of the appearances by transmitted light could be reproduced with photomacrographic images of simple, hand-drawn, periodic test-patterns.

Biology

CHAIRMAN: S. F. KERN

19. Electron Microscopy of Residual Chromosomes. A. R. T. DENUES, *Detroit Institute of Cancer Research, Detroit, Michigan.*—Residual chromosomes were prepared from chicken red blood cells by the general method of Mirsky and Ris with reiterated extractions with M NaCl. The matrix-containing supernatants from successive extractions were found by ultraviolet absorption spectrophotometry to absorb maximally near 2600A. Judging from this maximum absorption, the removal of matrix (desoxyribonucleohistone) from the chromosomes was incomplete in the first several extractions and approached completion only after about six washes when the volume-ratio of wash to whole chromosomes was of the order of 50. Preliminary electron micrographs revealed apparently coiled, filamentous submicroscopic structures that may intertwine and show crossbanding; these appeared cleanest after multiple extractions.

20. A New Fibrous Structure Obtained from Extracts of Collagenous Connective Tissue. J. GROSS,* J. H. HIGHBERGER, AND F. O. SCHMITT, *Biology Department, Massachusetts Institute of Technology, Cambridge, Massachusetts, and Research Division, United Shoe Machinery Corporation, Boston, Massachusetts.*—Various types of connective tissue including rat, calf and steer skin, rat tail tendon and the fish swim bladder were extracted with citrate buffer, $p\mathrm{H} \cong 4$. From the clear viscous extracts dialysis causes precipitation of a flocculent or fibrous material which was examined with the electron microscope. In addition to fibrils having a collagen-like axial periodicity a new type of fibrous structure was observed, namely fibrils with an axial structure, marked by dense thick bands, with repeating periods ranging from about 2000 to 3000A, depending on the source of material. Detailed intraperiod banding was frequently observed. Under certain conditions what appear to be individual segments of the long repeating structure were found. These appear as flat segments about 2000A long, frequently showing axial banding. The nature and origin of this new long-spacing material (which has not yet been observed in native connective tissue fibrils) will be discussed.

21. X-Ray and Electron Microscope Studies of Collagen. PAUL KAESBERG* AND M. M. SHURMAN, *Electron Microscope Laboratory and Department of Physics, University of Wisconsin, Madison, Wisconsin.*—A slit system and Geiger counter arrangement has been used to determine relative intensities for the first 30 orders of the x-ray diffraction pattern of a specimen of beef tendon collagen. These data are capable of yielding a Fourier series representation of the electron density distribution along the length of the fibers if the phases of the Fourier terms are known. The phases of the lower orders at least can be calculated from electron micrographs taken of the collagen specimens. Tentative phases have been assigned to all 30 orders and the resulting electron density distribution has been determined. Unstained specimens yield a relatively smooth rectangular density distribution. Phosphotungstic acid stained specimens give a characteristic fine structure similar to that obtained in electron micrographs of stained collagen.

22. The Ultrastructure of Nerve Axons Regenerated in Tissue Culture. EDUARDO DE ROBERTIS AND GIUSEPPE LEVI, *Department of Cell Ultrastructure. Institute of Biological Sciences, Montevideo, Uruguay,* to be presented by A. R. T. DENUES, *Detroit Institute of Cancer Research, Detroit, Michigan.*—Nerve axons regenerated from pieces of brain tissue of six to nine days old chick embryos were studied with the electron microscope. Axons from a few microns in diameter down to

sizes below the resolution of the optical microscope can be observed flattening down against the plastic film. Within the axon denser parallel lines along the axis indicate the limits between double-edged cylindrical structures similar to the neurotubules found in vertebrate peripheral nerves after fragmentation techniques. The diameter of these neurotubules varies between 300 to 400A, being smaller than the mean diameter of neurotubules from vertebrate nerves so far studied. In particularly favorable specimens the cross periodic banding with a period in the neighborhood of 600A can be observed. Observations are under way on nerve axons which are degenerating "in vitro" and on the growing processes of the axon where neurotubules are differentiating. The problem of artifacts in relation to the axon ultrastructure is discussed.

23. Uniformity of Lengths and Particle Weights of Tobacco Mosaic Virus. ROBLEY C. WILLIAMS,*† ROBERT C. BACKUS,† AND RUSSELL L. STEERE,‡ *University of Michigan, Ann Arbor, Michigan.*—Determinations have been made of the length distribution and of the particle weight of the particles of tobacco mosaic virus, by use of a modification of the spray-drop and counting technique developed by the authors. The lengths of the virus particles in spray-drop patterns containing less than 15 rods per pattern have been measured, and the results analyzed to ascertain the most probable length distribution in the original virus suspension. It is concluded that over 96 percent of the particles in the original suspension exist either in a monomeric length of 2980±100A, or in a dimeric multiple of this length. The particle weight of the virus has been determined by counting the number of virus particles in the electron micrographs of droplet patterns representing a known volume of a highly purified suspension, and by drying and weighing a known multiple of that volume. A value of 49×10^6 g/mole has been obtained for the weight of the monomeric particle.

† Now at the Virus Laboratory, University of California, Berkeley, California.
‡ Now at the Rockefeller Institute for Medical Research, New York City.

24. Ovomucin and the Egg White Inhibitor of Influenza Virus Hemagglutination. D. GORDON SHARP, *Duke University, Durham, North Carolina.*—From studies of the egg white inhibitor of influenza virus hemagglutination and of the enzymatic destruction of this inhibitor by active influenza virus, some physical data on ovomucin have been gathered. This mucoid fraction, called semipurified egg white inhibitor, has a viscosity increment of 273 to 453, and it is up to 60 times more active, as inhibitor, than egg white. It is complex, showing three components in electrophoresis with mobilities −3.5, −6.7, and −10 at *p*H 7.2 and 0.1 ionic strength. Only a single component is seen in the ultracentrifuge sedimenting at 40 *S* in fresh preparations of about 0.2 percent concentration but strongly dependent on concentration. Electron micrographs of this material (estimated molecular weight 10^7) show many threadlike particles, some as much as 1000A long. The widths are irregular, tapering to about 50A at the ends. These particles are responsible for most of the viscosity of dilute egg white and they are presumably responsible for its fibrous-supporting structure.

25. The Electron Micrography of the Wisconsin Pea-streak Virus. MARK A. STAHMANN, D. J. HAGEDORN, W. C. BURGER, AND PAUL KAESBERG,* *Departments of Biochemistry and Plant Pathology and the Electron Microscope Laboratory, University of Wisconsin, Madison, Wisconsin.*—Perfected Wales peas grown in the field were inoculated with the Wisconsin pea-streak virus. Concentration of the virus by differential centrifugation of the frozen extracts of diseased plants yielded highly infective pellets. Uranium shadowed electron micrographs prepared from pellets of diseased plants showed an abundance of long particles of unusual morphology. They were threadlike, filamentous and flexible and often several thousand mμ in length. Mounts made from pellets of healthy plants showed no such particles. Measurements of the particles indicated a width of about 12 mμ and lengths ranging from 150 to 5000 mμ. The microscope was calibrated by means of a metal-shadowed collodion replica of a standard diffraction grating. The pictures showed that the virus readily aggregates end-to-end and end-to-side. On standing in distilled water for several weeks, the virus underwent further aggregation to produce bundles of parallel strands which were often coiled about each other.

26. A Method of Obtaining Cultured Cells from Roller Tubes for Electron Microscopy. W. A. HOVIS,* G. O. GEY, AND F. B. BANG, *Department of Medicine and Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland.*—Roller tubes offer advantages over slides for cell cultivation over long periods. Electron microscope visualization of cells grown under these conditions is possible by coating the tubes with a Formvar plastic coating. Such membranes are not loosened by the flowing culture media. Cell strains have been studied and comparisons made between the results of this method and the cover-slip procedure. Virus infections of cells have been followed. A 0.25 percent solution of Formvar in ethylene dichloride leaves a thin film easily drained, and dried rapidly by aspiration with a water pump. The coated tubes are stored in a moist chamber and are good for at least a month with good membranes. Many cells grown in such tubes are thinly spread out, as good as those on glass. The cell-coated membrane can be removed at any culture period, after fixation, by breaking open the tubes, and harvesting in the usual manner for E. M.

27. Production of External and Internal Filamentous Structures in a Non-Ciliated Strain of "Bacillus Subtilis." VALENTIN BONIFAS, EDOUARD KELLENBERGER, AND GEORGES H. WERNER,*† *Electron Microscope Laboratory of the University of Geneva, Switzerland.*—Filamentous networks including in their meshes some granules of cytoplasmic material are observed as remains in many cells of this strain of B. subtilis, that have been ruptured by a number of physico-chemical processes. Production of external filamentous structures is an effect of suspension of the cells in distilled water, in sodium bicarbonate solutions as well as of their contact with antiseptics in the culture medium, and this production is regularly observed during sporulation. This phenomenon affects only a rather low percentage of the cells. Some characteristics of these filaments are described. The significance of this observation is discussed in correlation with previous findings of R. W. G. Wyckoff, and W. van Iterson.

† Present address: Squibb Institute for Medical Research and Public Health Research Institute of the City of New York.

28. Electronmicrography of Treponemas from Cases of Yaws, Pinta and the So-Called Cuban Form of Pinta. JUAN J. ANGULO, JOHN H. L. WATSON,* C. COURTNEY WEDDERBURN, FRANCISCO LEÓN-BLANCO, AND GERARDO VARELA, *Department of Experimental Pathology, University of Havana School of Medicine, Havana, Cuba; Edsel B. Ford Institute for Medical Research, Detroit, Michigan; Government Medical Serivce, Kingston, Jamaica, B. W. I.; Commission for the Prophylaxis of Syphilis, Leprosy and Cutaneous Diseases, Havana, Cuba; and Institute of Tropical Diseases, Mexico City, Mexico.*—The morphologic and structural findings made with the electron microscope in unshadowed and in shadow-cast treponemas from typical cases of yaws, pinta and the so-called Cuban form of pinta are presented. Thirty *T. pertenue*, nine *T. carateum* specimens as well as eighty-nine treponema specimens from a case of the Cuban form of pinta constitute the material studied. No differences in morphology or structure were detected among these species or with reference to *T. pallidum* except for the lack of flagellae shown by *T. pertenue* which may have

been an artifact. Measurements of lengths and diameters were made on the treponemas. Several examples of pseudostructures were met, some of which had a bearing upon the significance of "spore-like bodies" and "end bodies."

29. On the "Fixation" of Electron Microscope Specimens by Electron Bombardment. J. HILLIER, S. MUDD, AND A. G. SMITH, presented by S. G. ELLIS, *RCA Laboratories, Princeton, New Jersey. RCA Laboratories, Princeton, New Jersey and the University of Pennsylvania, Philadelphia, Pennsylvania.*—A number of the effects of the bombardment of specimens in the electron microscope have been reported. Small, osmium-fixed bacteria could be hydrolyzed after being subjected to the vacuum of the instrument but not after the minimum bombardment necessary for making a micrograph. A two-second exposure of a collodion supporting membrane to a bombardment of 0.01 amp./cm² made it insoluble in amyl acetate, acetone and ether-alcohol and raised its melting point above 600°C. It is suggested that such a bombarded membrane will be useful for mounting materials from solvents. A collodion and paraffin embedded section subjected to a bombardment of only 0.001 amp/cm² for 3 or 4 seconds showed no visible change in the electron microscope, but was changed chemically such that the embedding material could no longer be removed by any of the common solvents. The stability of the specimens after bombardment is considered to be the result of a binding between adjacent molecules. The theory that considers the stable specimens to be the carbonized remains of the original is not supported by this work at low intensities.

30. Recent Electron Microscopic Work at Leeds. W. T. ASTBURY, *Chairman, British Electron Microscopy Group, Department of Biomolecular Structure, University of Leeds, Leeds, England.*

31. Light and Electron Microscopical Observations of Nerve Fibers Treated with Chemical and Bacterial Enzymes. P. LEPINE, *Pasteur Institut, Paris, France.*

32. Electron Microscope Observations on Lysed and Ground Staphylococcus Aureus and Micrococcus Lysodeikticus. S. F. KERN, R. A. KERN,* O. K. BEHRENS, AND M. J. KINGKADE, *Lilly Research Laboratories, Indianapolis, Indiana.*—Staphylococcus aureus and Micrococcus lysodeikticus are affected differently by lysozyme. Electron micrographs made of Micrococcus lysodeikticus after various times of action of lysozyme show a sequence of events. The flattening of the cell is followed by the rupture of the cell membrane, the dissolution of the cell contents, and finally the dissolution of the cell membrane. Cell bodies in the Micrococcus lysodeikticus are not attacked by the lysozyme. Lysozyme on Staphylococcus aureus dissolves a portion of the cell but other constituents are apparently not attacked. Electron micrographs of ground Micrococcus lysodeikticus confirm the presence of cell bodies and the cell membrane. Grinding also reveals a definite cell membrane on Staphylococcus aureus.

33. Electron Microscopical Study of Crystalline Macromolecular Diethyl Germanium Oxide. E. G. ROCHOW, *Harvard University, Cambridge, Massachusetts,* AND T. G. ROCHOW,* *American Cyanamid Company, Stamford, Connecticut.*—$(C_2H_5)_2GeO$ can be made to polymerize to form macromolecules which are electron microscopic in size, a high proportion of which are in the range of sizes and shapes of viruses. Likewise, the "germanione" macromolecules form primitive crystals. Electron micrographs will illustrate these observations.

34. Electron Microscopy in Japan. NOBORU HIGASHI, M.D., Present address, *Research Fellow, Department of Infectious Diseases, School of Medicine, University of California, Los Angeles;* Permanent address, *University of Kyoto, Japan*

Microtomy and Miscellaneous

CHAIRMAN: JOHN J. KELSCH

35. Techniques for Use in the Preparation of Thin Sections. SANFORD B. NEWMAN AND EMIL BORYSKO,* *National Bureau of Standards, Washington, D. C.*—Techniques are described which surmount some of the difficulties encountered in the preparation and study of thin sections. A method for handling partially polymerized methacrylate, used to prevent tissue damage in the embedding process, is described in which a drawn-out test tube serves as a polymerization vessel and dispenser. A technique has been developed for centering specific areas of a section on weakly magnetic stainless steel mounting screens by means of an electromagnet held in a micromanipulator. A technique is described for digesting sections with enzymes, resulting in the differential removal of cellular components. The value of this technique in the study of tissues is discussed. A density-volume method for determining the thickness of sections is presented. The possibility of using sapphire, tungsten carbide, beryllium copper and phosphor bronze for microtome knives is discussed.

36. Electron Microscopy of the Kidney. DANIEL C. PEASE* AND RICHARD F. BAKER, *Departments of Anatomy and Experimental Medicine, School of Medicine, University of Southern California, Los Angeles, California.*—Ultrathin sections of rat kidney have been studied with the electron microscope. Technique is discussed. It has been found that the brush border of the proximal tubule consists of myriads of minute processes extending into the lumen which presumably increase vastly the surface area available for resorption. The mitochondria of the proximal tubule presented a varying picture depending upon their state of activity. Two somewhat independent processes were recognized, one leading to the formation of large vacuoles, the other to diffuse liberation of mitochondrial material into the cytoplasm. Glomerular capillaries were found not to have a continuous endothelium, although the epithelial layer may have been complete. The basement membrane was peculiarly specialized with a ribbed surface, possibly to afford support. The extraglomerular capillary system of the cortex, and the venous capillaries of the medulla also lacked a continuous endothelium. Other matters considered include the special morphology of different parts of the nephron, connective tissue elements, unusual features of the arterial capillaries of the medulla and of the interlobar arteries, the role of cytoplasmic elements in the orientation of mitochondria, and a general discussion of basement membranes.

37. Further Developments in the Electron Microscopic Study of Tissue Sections. JAMES HILLIER, *RCA Laboratories, Princeton, New Jersey and Sloan-Kettering Institute for Cancer Research, New York, New York,* to be presented by CECILE CANNAN, *Sloan-Kettering Institute, New York City.*—A number of methods of dissolving the embedding were studied, but none was found in which the distortions introduced could be reduced to a level below the thickness of the section. Judicious bombardment by the electron beam removes most of the embedding and appears to preserve structures to the 200A level. A new objective lens pole piece was designed specially for the micrography of sections. This lens operates at a focal length of 12 mm. It includes a diaphragm which is placed at the image side focal point to give large bright field with angular apertures as low as 2×10^{-3} rad. With the pole piece removed, the aperture can be centered or shifted along the axis of the lens. Some examples of the results obtained are shown.

38. Electron Microscopy of Thin-Sectioned Spirostomum. HAROLD E. FINLEY, *Department of Zoology, Howard University, Washington, D. C.*—The National Bureau of Standards microsectioning method was used in a study of the ciliated protozoan

Spirostomum ambiguum. Fine details of many parts of thin-sectioned organisms were revealed by the electron microscope. Among the fine details of morphology disclosed, the following were prominent: Compact bundles of fibrils in locomotory cilia; basic architectural units of the ectoplasmic matrix (fibrils); basal bodies of membranellae and interciliary fibrils; particles in the macronucleus; and ingested organisms in food vacuoles. These fine structures will be illustrated and discussed.

39. A Method of Preparing Thin Sections of Grease for the Electron Microscope. R. BAKER,* O. WARREN, H. COFFER, R. VOLD, AND D. PEASE, *University of Southern California, Los Angeles, California.*—A technique which allows 0.1-micron sections of a simple two-component grease to be cut is described. The grease used is 25 percent calcium stearate monohydrate and 75 percent cetane. After freezing with dry ice, a slab 20 by 5 by 3 mm is cut and mounted in a modified Spencer microtome. A ribbon of sections is collected in a cardboard trough containing water which is sealed to the knife edge. The block is kept frozen during cutting, and it is necessary also to chill the knife and collecting fluid below the freezing point of cetane (18°C). The microtome is motor-driven and optimum cutting speed is approximately 150 sections per minute. Cetane is extracted from the sections with benzene, which is then removed either by air-drying or freeze-drying. A simple mounting procedure is described.

40. On the Sharpening of Microtome Knives for Ultrathin Sectioning. JAMES HILLIER, *RCA Laboratories, Princeton, New Jersey and Sloan-Kettering Institute for Cancer Research, New York, New York*, to be presented by CECILE CANNAN, *Sloan-Kettering Institute, New York City.*—Microtome knives which have become dull cutting ultrathin sections show little, if any, visible change under the high power light microscope. This indicated that only a minimum amount of repolishing should be necessary to regain an edge. A study of conventional methods showed that most of the sharpening time involved the unnecessary removal of material in order to fit the edge to the sharpening means. A short permanently attached sharpening back was designed, which eliminated the two most important sources of error: change of sharpening angle and lack of parallelism between the back and the edge. With this back, it is possible to obtain ideal edges in as little as five minutes of sharpening. The technique of preparing the blade for this technique and the sharpening procedure will be described in some detail. Prepared glass plates are used as sharpening surfaces. Levigated alumina is used for grinding and Linde A5175 Metal polishing powder is used for polishing. The edges are never stropped.

41. A Simplified Microtome for Electron Microscopy. G. G. COCKS AND C. M. SCHWARTZ,* *Battelle Memorial Institute, Columbus, Ohio.*—A microtome has been built for cutting sections suitable for electron microscopy. The essential feature of the design is the elimination of bearings and ways ordinarily required for control of the traverse of the specimen with respect to the knife. The specimen is fixed to the end of a long cantilever which when deflected carries the specimen across the knife. The bending stresses are well within the elastic limit of the cantilever material so that the sample returns precisely to its original position. Provision for torsional rigidity and stability is incorporated in the design. No precision machining or heat treatment of parts is required, and the cost is low. Specimen feed is by thermal expansion, at present.

42. A Microtome Specimen Holder Advanced by Thermal Expansion. M. EDEN,* A. W. PRATT, and H. KAHLER, *National Cancer Institute, Bethesda, Maryland.*—A microtome specimen holder that advances the specimen by thermal expansion is described. The holder is heated by an electrical current, the value of which can be regulated to give any desired rate of heating and expansion. The rate of expansion does not diminish by more than 20 percent for a range of over three microns. The rate of expansion is reproducible to within three percent over the same range. A quantitative comparison has been made between the heating and expansion rates of the CO_2 precooled instrument and the electrically heated instrument.

43. Electron Microscopical Determination of Macromolecular Weights in Resins. T. G. ROCHOW,* M. C. BOTTY, AND F. G. ROWE, *Research Laboratories, American Cyanamid Company, Stamford, Connecticut.*—Using the example of a high polymer of unit density with its macromolecules retracted as spheres, a diameter of 50A corresponds to a molecular weight (M) of 40,000 and 140A corresponds to $M=1{,}000{,}000$. Such dimensions are electron microscopical and the corresponding molecular weights (or higher ones) are dominant in practical plastics and extensomers. The molecular weights are easily computed from the density (d) and Avogadro's number (N), thus: $M=(\pi/6)Nd^3=3.15d^3\times10^{23}$. Two other methods of calculation are independent of N. Such calculations give distributions and averages of molecular weight. The electron micrographs also depict rather than infer the shapes of the macromolecules. For soluble resins, electron microscopical data will be compared with those obtained by conventional methods. Macromolecular data may be uniquely obtained on *insoluble* resins by replicating the fracture surfaces of the embrittled resins.

44. The Structure of Certain Aircraft Lubricating Greases. J. I. WITTEBORT* AND BERNARD RUBIN, *Air Matériel Command, Wright-Patterson Air Force Base, Dayton, Ohio.*—Techniques of electron microscopy have been found useful in determining the structure and morphological differences in the solid phase of several aircraft lubricating greases. Greases having a solid phase consisting of the stearate soaps of sodium, calcium, lithium, barium, strontium, and aluminum have been studied. With the exception of aluminum stearate grease each of the various stearate greases has a characteristic fibrous structure which goes to form a matted felt-like mass entrapping the oil phase in its interstices. The fiber size of lithium stearate greases gave evidence of being influenced by the chemical nature of the oil phase. Lithium greases having the oil phase composed of petroleum, diester, silicone and silicone-diester types were investigated. Electron microscopic methods offer a rapid and effective means of identifying the solid phase of lubricating greases and studying the effects of various chemical and mechanical treatments on lubricating greases.

45. The Effect of Heating Evaporated Metal Films. R. S. SENNETT, *University of Toronto, Toronto, Ontario.*—Thin films of a number of metals have been heated to different temperatures and allowed to cool in vacuum and the changes in structure and properties have been studied. In the case of thin silver films the aggregates became more regular with greater separations between them while thicker films which appeared continuous before heating became aggregated. This caused marked changes in the optical and electrical properties of the films. The characteristic maxima in the variation of the absorption with thickness became broader and moved to greater thicknesses. Also after heating the thickness required to show measurable electrical conductivity was greater. Similar changes on a smaller scale were observed in gold, chromium, and aluminum films.

46. Electron Microscopy of Cotton Fiber Constituents. PAULINE E. HOLBERT* AND WANDA K. FARR, *Celanese Corporation of America, Summit, New Jersey.*—Dissection of the cotton fiber into its sheath-like primary wall and fibrillar secondary lamellae has been followed by identification of the constituents cellulose, pectic material, protein, and wax by means of microchemical, polarized light, and refractive index

techniques. Electron micrographs of the untreated, noncellulosic, primary wall reveal no structure at magnifications of 5000 and 15,000 diameters. The primary wall substance is comparatively insensitive to the electron beam (bombardment). Dissected fibrils from the secondary wall have a much greater density than the primary wall material and are extremely sensitive to the electron beam. Until suitable mounts were made which provided protection, the constituent cellulose particles in these fibrils suffered gross expansion and distortion of their internal structure, although their original ellipsoid shape was maintained. In the final preparations the particles from both young and old fibers have the dimensions and physical properties of those described in previous publications.

47. Electron Microscopy of Radiation Polymerization and Condensation Products from Controlled Experiments. JOHN H. L. WATSON* AND MARCEL VANPEE, *Edsel B. Ford Institute for Medical Research, Detroit, Michigan.*—Solid and liquid residues form when certain gases are subjected to radiations. As a possible new approach to the study of the reaction mechanisms of radiation chemical processes these residues formed under controlled conditions are investigated by the methods of electron microscopy, some of the physical effects of radiation upon matter being observable by these means. In the present experiments the gases CO, C_2H_2, and CH_2CHCl have been bombarded individually under a variety of conditions with alpha-particles from radon and the residues collected directly upon prepared specimen screens within the chamber. The characteristics of the deposits are studied with varying exposure times and varying distances from the alpha-particle source. Information is made available on the particle size and shape of the solid fraction and the liquid droplets; on the relative abundance of solid and liquid phases and on the morphology of the whole. Inferences, regarding the mechanism of the reactions, may be drawn from these observations. Crystalline forms of the C_2H_2 and CH_2CHCL were observed clearly in the residues.